



HAL
open science

Transcription factors **KANADI 1**, **MYB DOMAIN PROTEIN 44**, and **PHYTOCHROME INTERACTING FACTOR 4** regulate long intergenic noncoding RNAs expressed in *Arabidopsis* roots

Li Liu, Michel Heidecker, Thomas Depuydt, Nicolas Manosalva Perez, Martin Crespi, Thomas Blein, Klaas Vandepoele

► To cite this version:

Li Liu, Michel Heidecker, Thomas Depuydt, Nicolas Manosalva Perez, Martin Crespi, et al.. Transcription factors **KANADI 1**, **MYB DOMAIN PROTEIN 44**, and **PHYTOCHROME INTERACTING FACTOR 4** regulate long intergenic noncoding RNAs expressed in *Arabidopsis* roots. *Plant Physiology*, 2023, 193 (3), pp.1933-1953. 10.1093/plphys/kiad360 . hal-04299815

HAL Id: hal-04299815

<https://hal.science/hal-04299815>

Submitted on 22 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Transcription factors KANADI 1, MYB DOMAIN**
2 **PROTEIN 44, and PHYTOCHROME INTERACTING**
3 **FACTOR 4 regulate long intergenic noncoding RNAs**
4 **expressed in Arabidopsis roots**

5 Li Liu^{1,2}, Michel Heidecker^{3,4}, Thomas Depuydt^{1,2}, Nicolas Manosalva Perez^{1,2},
6 Martin Crespi^{3,4}, Thomas Blein^{3,4*}, Klaas Vandepoele^{1,2,5*}

7

8 (1) Ghent University, Department of Plant Biotechnology and Bioinformatics,
9 Technologiepark 71, 9052 Ghent, Belgium

10 (2) VIB Center for Plant Systems Biology, Technologiepark 71, 9052 Ghent,
11 Belgium

12 (3) Université Paris-Saclay, CNRS, INRAE, Université Evry, Institute of Plant
13 Sciences Paris-Saclay (IPS2), 91190 Gif-sur-Yvette, France

14 (4) Université Paris Cité, CNRS, INRAE, Institute of Plant Sciences Paris-Saclay
15 (IPS2), 91190 Gif-sur-Yvette, France

16 (5) Bioinformatics Institute Ghent, Ghent University, Technologiepark 71, 9052
17 Ghent, Belgium

18

19 * Shared last author

20 Corresponding author: Klaas Vandepoele, klaas.vandepoele@psb.vib-ugent.be

21

22 The author responsible for distribution of materials integral to the findings
23 presented in this article in accordance with the policy described in the
24 Instructions for Authors ([https://academic.oup.com/plphys/pages/General-](https://academic.oup.com/plphys/pages/General-Instructions)
25 [Instructions](https://academic.oup.com/plphys/pages/General-Instructions)) is _.

26 One-sentence summary:

27

28 Running title: Regulatory annotation of Arabidopsis root lincRNAs

29

30 **ABSTRACT**

31 Thousands of long intergenic noncoding RNAs (lincRNAs) have been identified in
32 plant genomes. While some lincRNAs have been characterized as important
33 regulators in different biological processes, little is known about the
34 transcriptional regulation for most plant lincRNAs. Through the integration of
35 eight annotation resources, we defined 6,599 high-confidence lincRNA loci in
36 *Arabidopsis* (*Arabidopsis thaliana*). For lincRNAs belonging to different
37 evolutionary age categories, we identified major differences in sequence and
38 chromatin features, as well as in the level of conservation and purifying selection
39 acting during evolution. Spatiotemporal gene expression profiles combined with
40 transcription factor (TF) chromatin immunoprecipitation data were used to
41 construct a TF-lincRNA regulatory network containing 2,659 lincRNAs and
42 15,686 interactions. We found that properties characterizing lincRNA expression,
43 conservation and regulation differ between plants and animals. Experimental
44 validation confirmed the role of three TFs, KAN1, MYB44, and PIF4, as key
45 regulators controlling root-specific lincRNA expression, demonstrating the
46 predictive power of our network. Furthermore, we identified 58 lincRNAs,
47 regulated by these TFs, showing strong root cell-type specific expression or
48 chromatin accessibility, which are linked with GWAS genetic associations related
49 to root system development and growth. The multi-level genome-wide
50 characterization covering chromatin state information, promoter conservation,
51 and ChIP-based TF binding, for all detectable lincRNAs across 769 expression
52 samples, permits rapidly defining the biological context and relevance of
53 *Arabidopsis* lincRNAs through regulatory networks.

54

55 One-sentence summary:

56 A multi-level *Arabidopsis* gene regulatory network identifies regulators controlling
57 root-specific lincRNA expression, offering a promising strategy to identify
58 lincRNAs involved in plant biology.

59

60

61 **Introduction**

62 Genomes are widely transcribed and produce thousands of long non-coding
63 RNAs (lncRNAs), which are an abundant class of transcripts longer than 200
64 nucleotides with low protein coding capacity (Wu et al., 2017). LncRNAs are
65 generally transcribed by RNA polymerase II (Pol II), and are processed in a
66 similar way as mRNAs, with capping, splicing, and polyadenylation. LncRNAs
67 modulate gene expression through a wide range of mechanisms, including
68 chromatin structure remodeling, transcription regulation in cis/trans, fine-tuning of
69 miRNA activity, alternative splicing (AS) regulation, and the control of mRNA
70 stability and translation (Sanchita et al., 2020; Bhogireddy et al., 2021; Lucero et
71 al., 2021). One class of lncRNAs is the primary mRNAs containing miRNA
72 precursors or pri-miRNAs. These are rapidly and generally processed into
73 miRNAs (Jones-Rhoades et al., 2006; Shafiq et al., 2016). However, the large
74 majority of lncRNAs are able to act without being further processed such as the
75 lncRNAs controlling the epigenetic regulation of *Flowering Locus C (FLC)* gene
76 expression and mediating plant vernalization, i.e., *COOLAIR*, *COLDAIR* and
77 *COLDWRAP* (Liu et al., 2010; Heo and Sung, 2011; Kim and Sung, 2017). A
78 subgroup of lncRNAs derived from intergenic regions is defined as long
79 intergenic non-coding RNAs (lincRNAs) and have been identified in a wide range
80 of eukaryotes including model and non-model plant species (Wu et al., 2017;
81 Chen et al., 2021). In contrast to antisense lncRNAs, whose sequence evolution
82 is constrained by the overlapping coding genes, the transcription and evolution of
83 lincRNAs are independent of the surrounding genes.

84 While the low expression levels and tissue-specific expression patterns of
85 lincRNAs in plants initially raised concerns (Liu et al., 2012; Bu et al., 2015),
86 increasing experimental evidence supports the functional activity of lincRNAs. In
87 plants, few lincRNAs have been experimentally validated (Chen et al., 2021),
88 showing their involvement in various biological contexts such as in regulating
89 flowering time (Chen et al., 2020) and root growth and development (Roule et al.,
90 2022a), or influencing germination (Kramer et al., 2022). For example, the
91 *Arabidopsis (Arabidopsis thaliana) FLINC* lincRNA has been reported to regulate

92 ambient temperature-mediated flowering time (Severing et al., 2018). Arabidopsis
93 lateral root development is regulated by the *Alternative Splicing COmpetitor*
94 (*ASCO*) lincRNA, which modulates AS by interacting with the multiple splicing
95 factors (Bardou et al., 2014; Rigo et al., 2020), and the *AUXIN-REGULATED*
96 *PROMOTER LOOP* (*APOLO*) lincRNA, which influences the local chromatin
97 conformation and the activity of several Auxin-Responsive genes (Ariel et al.,
98 2020). Plants lacking *CONSERVED INBRASSICA RAPA1* (*IncCOBRA1*) were
99 found to show a delayed germination and were overall smaller than wild-type
100 plants (Kramer et al., 2022). Many lincRNAs are differentially expressed in
101 various stress responses, including drought (Qi et al., 2013; Shuai et al., 2014; Li
102 et al., 2017; Qin et al., 2017), cold (Li et al., 2017; Zhao et al., 2018; Shea et al.,
103 2019), salinity (Deng et al., 2018; Fukuda et al., 2019), and nutrient deficiency
104 (Fukuda et al., 2019), implying that lincRNAs may be involved in plant stress
105 resilience (Jha et al., 2020). Interestingly, some of the confirmed functional
106 lincRNAs interact with transcription factors (TFs) to activate or repress the
107 expression of associated genes, such as *APOLO* that interacts with *WRKY42* to
108 form a regulatory hub that controls the activity of *RHD6* and promotes the
109 expansion of root hair cell at low temperatures (Moison et al., 2021; Pacheco et
110 al., 2021).

111 Recently, a lot of attention has been placed on the evolutionary conservation of
112 lincRNAs, which is generally associated with functionality (Ulitsky, 2016;
113 Szczesniak et al., 2021). The conservation of noncoding transcripts can be
114 considered at the level of the primary sequence, position, splice sites, RNA
115 structure, and transcriptional level (Ulitsky, 2016; Szczesniak et al., 2021).
116 However, most lincRNA sequences undergo rapid evolution and are poorly
117 conserved (Ransohoff et al., 2018). In a study by Wang et al., only 5% of 117 rice
118 (*Oriza sativa*) lincRNAs had sequence similarity to maize (*Zea mays*) lincRNAs. It
119 was also found that the positional conservation of lincRNAs was much higher
120 than their sequence conservation (Wang et al., 2015). Nelson and co-workers
121 reported that 22% of 1180 Arabidopsis lincRNA loci were conserved in 10
122 Brassicaceae genomes (Nelson et al., 2016). These conserved lincRNAs

123 exhibited higher expression levels, stress-responsiveness and their gene body
124 overlapped with conserved noncoding sequences (CNSs), suggesting a role of
125 their conserved sequence in a genomic context (Nelson et al., 2016).
126 While different studies have reported on the identification and expression of
127 lincRNAs in plants (Wang et al., 2015; Nelson et al., 2016; Ke et al., 2019; He et
128 al., 2021), a comprehensive overview of the different genomic features
129 contributing to the expression, regulation and evolutionary conservation of plant
130 lincRNAs is missing. How lincRNAs are embedded in transcriptional networks
131 controlling different biological processes remains largely unknown. Furthermore,
132 prioritizing lincRNAs for downstream functional analysis is not straightforward
133 without knowing the regulatory network where they are involved in. Here, we
134 integrated different *Arabidopsis* lincRNA gene annotations and explored various
135 functional genomics datasets to characterize lincRNA expression in a biologically
136 relevant context. We leveraged large-scale expression datasets and protein-DNA
137 interaction data to study the molecular networks controlling lincRNA gene activity.
138 Combined with evolutionary conservation analysis, we explored the global
139 transcriptional regulatory properties of different evolutionary age categories and,
140 through regulatory network analysis, identified specific TFs controlling lincRNA
141 regulation in roots.

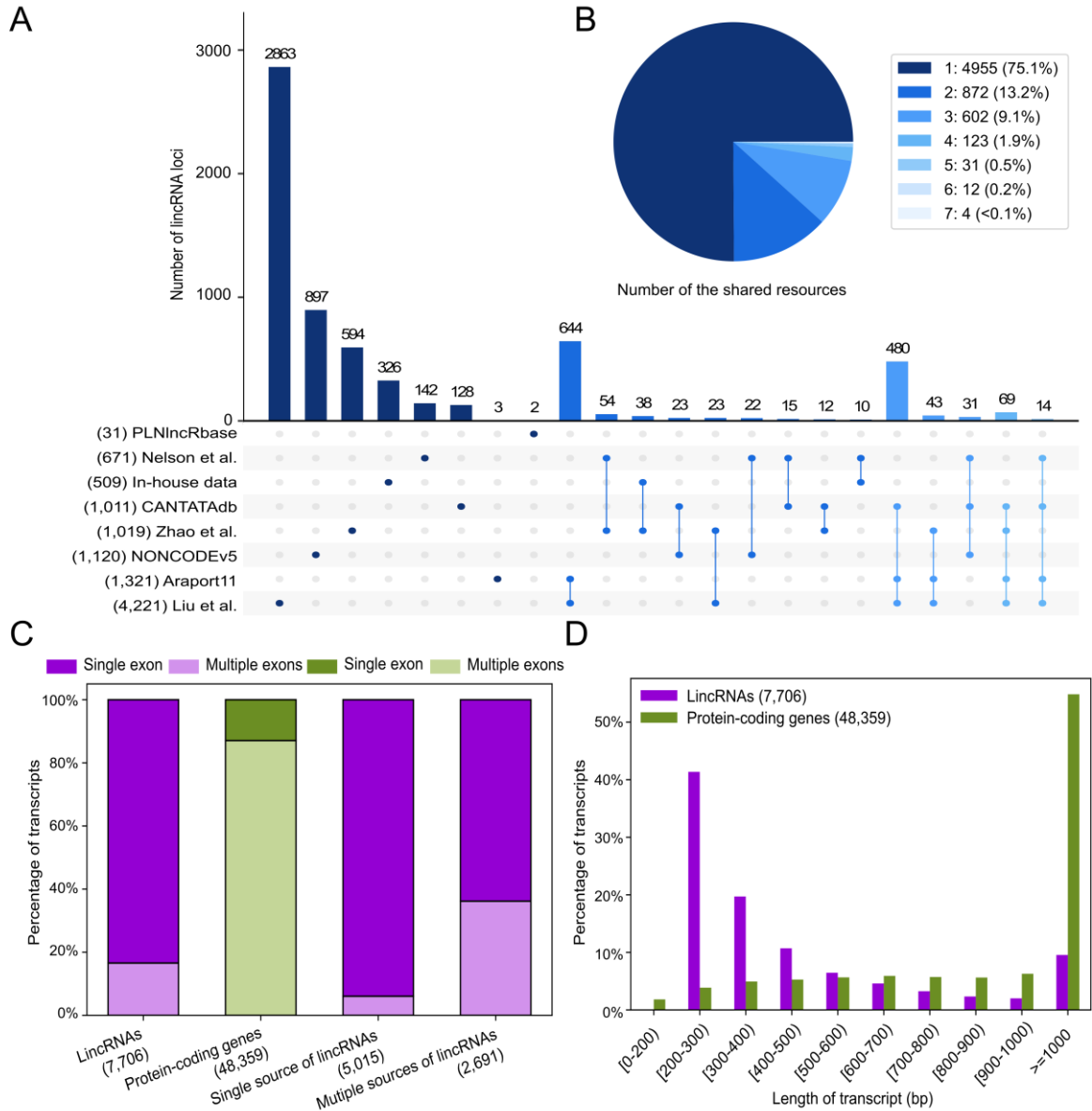
142
143

144 **RESULTS**

145 **Integration of lincRNA annotations in *Arabidopsis***

146 A substantial number of lincRNA transcripts in *Arabidopsis* have been identified
147 and several publicly available resources for the annotation of lincRNAs have been
148 developed (Jha et al., 2020). In contrast to antisense lincRNAs, which generally
149 regulate the overlapping coding gene, much less is known about the potential
150 targets of lincRNAs, so we focus our study on these transcripts. Indeed,
151 lincRNAs transcription and evolution are independent of the surrounding genes,
152 in contrast to antisense lincRNAs that are constrained by the coding genes they
153 overlap with. To integrate and unify previously identified lincRNAs, annotations

154 based on transcriptome information from 10 different tissues and various
155 environmental conditions were collected from different resources including
156 Araport11 (Cheng et al., 2017), CANTATAdb (Szczesniak et al., 2016),
157 NONCODEv5 (Fang et al., 2018), PLNlncRbase (Xuan et al., 2015), key lncRNA
158 research articles (Liu et al., 2012; Nelson et al., 2017; Zhao et al., 2018), and
159 predictions based on root related stranded RNA-seq (Materials and Methods).
160 Next, a pipeline was designed to define a unified set of high-confidence lincRNAs
161 by discarding transcripts with length below 200 bp, removing transcripts that
162 overlapped with protein-coding genes (antisense lncRNAs), re-evaluating the
163 coding potential of the transcripts, and merging the remaining transcripts from
164 various resources (see Materials and Methods, Supplemental Figure S1A-B). In
165 total, we identified 7,706 lincRNA transcripts covering 6,599 high-confidence
166 lincRNA loci (see Supplemental Data Set S1). To explore the overlap of this high-
167 confidence lincRNA gene set with the individual annotations, we assessed the
168 overlap between the different resources (Figure 1A). A total of 4,955 (75.1%)
169 lincRNA loci were supported by only one resource and the remaining 1,644
170 (24.9%) lincRNA loci were derived from two or more resources (Figure 1B).
171 Araport11 contained the highest number of shared loci and Liu et al. (2012)
172 contained the highest number of unique loci. Next, the genomic features of the
173 high-confidence lincRNA transcripts were compared to those of protein-coding
174 transcripts. 6,428 (83%) lincRNA transcripts and 6,250 (13%) protein-coding
175 transcripts contained single exons, while 1,278 (17%) lincRNA transcripts and
176 42,109 (87%) protein-coding transcripts contained multiple exons. Furthermore, a
177 higher frequency of multi-exon transcripts was found in lincRNAs supported by
178 two or more resources (974, 36%) than in those supported by a unique resource
179 (304, 6%) (Figure 1C). The transcript length distribution for lincRNAs showed a
180 U-shape curve with the majority of transcripts being 200-300bp long (Figure 1D).



181

182 **Figure 1. Overlap and gene features of Arabidopsis lincRNA annotations. (A)**

183 Upset plot showing the intersection of lincRNA annotation in the eight resources.

184 Each row represents a resource, reporting in parenthesis its total number of

185 lincRNA transcripts before merging. LincRNA annotations unique to a single

186 resource are represented as a single circle while circles connected by lines

187 represent the intersection of lincRNA loci shared between various resources. The

188 bar chart indicates the number of unique lincRNA loci and intersectional lincRNA

189 loci, displaying only intersections that contain at least ten lincRNA

190 complex overlapping patterns are not shown. **(B)** The pie chart shows the

191 proportion of lincRNA loci supported by one or more resources. **(C)** The
192 distribution of exon number for all lincRNA transcripts (purple), protein-coding
193 transcripts (green), transcripts of lincRNAs supported by single resource (purple)
194 and multiple resources (purple). Single exon and multiple exons are shown in
195 dark and light colors, respectively. **(D)** The distribution of transcript length for
196 lincRNAs (purple) and protein-coding genes (green).

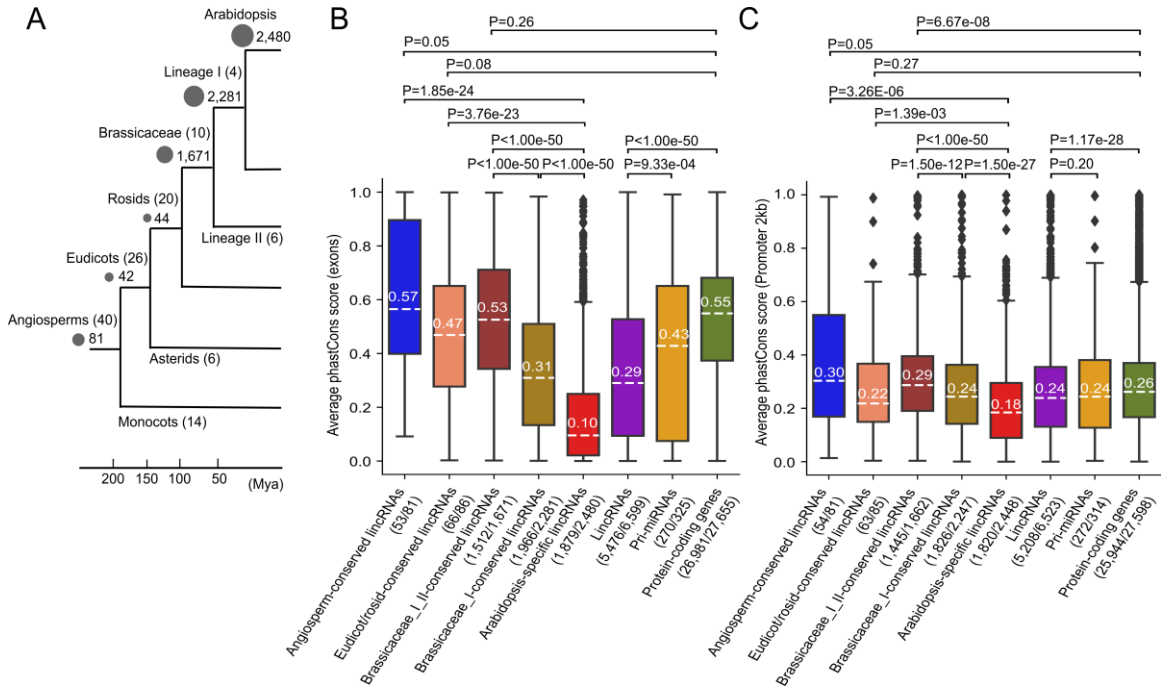
197

198

199 **Contrasting patterns of sequence conservation for lincRNAs belonging to** 200 **different evolutionary age categories**

201 To assess the evolutionary conservation of Arabidopsis lincRNAs within flowering
202 plants, DNA sequence similarity searches were performed by comparing our set
203 of lincRNAs with the genomes of 40 plant species (see Materials and Methods,
204 Supplemental Table S1). Among the 6,599 lincRNAs, 2,480 lincRNAs were
205 restricted to Arabidopsis and named Arabidopsis-specific lincRNAs. The other
206 lincRNAs were classified into four evolutionary age categories according to the
207 presence of homologs in closely and more distantly related species (Figure 2A).
208 We found 81 lincRNAs with at least one homolog in eudicots and in monocots
209 and therefore conserved during 180 million years (MY) of evolution (Beilstein et
210 al., 2010; Zhang et al., 2020), defined as angiosperm-conserved lincRNAs. Forty-
211 two lincRNAs were conserved in eudicots with at least one homolog in rosids and
212 asterids, but without homologs in monocots. Similarly, 44 lincRNAs were
213 identified only having homologs in rosids species, outside the Brassicaceae
214 family. As the lincRNAs conserved in eudicots and rosids showed highly similar
215 conservation patterns for exon and promoter ($P < 0.264$, Mann–Whitney U test),
216 we combined these genes in one category, called Eudicot/rosid-conserved
217 lincRNAs (86 genes). 1,671 Brassicaceae_I_II-conserved lincRNAs were present
218 in the common ancestor of Brassicaceae lineages I and II, without homologs
219 outside the Brassicaceae. Lastly, 2,281 lincRNAs were identified only having
220 homologs in the Brassicaceae I lineage (Brassicaceae_I-conserved lincRNAs).

221 The monotonous decrease in the number of lincRNAs for the older age
 222 categories suggests there is a continuous birth of lincRNA loci at the species
 223 level, with only a small fraction showing deep conservation in other plant families
 224 or orders.



225
 226 **Figure 2. Sequence conservation analysis for lincRNAs of different**
 227 **evolutionary age categories. (A)** Simplified species tree reporting the number
 228 of lincRNAs found for the different age categories, which are also indicated by
 229 the grey circle sizes. Numbers in parenthesis report the number of genomes
 230 included per clade to assess sequence similarity and define a lincRNA's age
 231 category. Boxplot showing the average phastCons score for **(B)** exons and **(C)**
 232 promoter regions (2kb upstream of transcription start site) of different lincRNA
 233 age categories and gene types (lincRNAs, pri-miRNAs, protein-coding genes).
 234 The numbers in parentheses report the number of exons and promoters with at
 235 least 50% of informative nucleotides over the total number of gene bodies and
 236 promoters in that category, respectively. PhastCons score ranges from 0 (not
 237 under selection) to 1 (strong negative selection). P-values for pairwise Mann-
 238 Whitney *U* test are shown using the horizontal lines connecting the series and
 239 were corrected for multiple testing using the Benjamini-Hochberg procedure.

240

241

242 Apart from detecting lincRNA homologs, we also evaluated evolutionary selection
243 of lincRNA loci, pri-miRNAs, where globally the 21-24 miRNA sequence is the
244 only conserved region at the nucleotide level, and protein-coding genes using
245 phastCons scores (Siepel et al., 2005). This score reports the probability for each
246 nucleotide to evolve neutrally or under negative, or purifying, selection (low or
247 high phastCons score, respectively). In contrast to the age categories, which are
248 based on finding similar homologous sequences in other plant genomes and do
249 not give information about the mode of selection, phastCons works by fitting a
250 two-state hidden Markov model to a genome-wide multiple sequence alignment
251 and predicting, based on the pattern of nucleotide substitutions, conserved
252 elements representing sites under purifying selection (Siepel et al., 2005).
253 Consequently, whereas the age categories give an indication of the emergence
254 of an individual gene within the green plant lineage, the phastCons scores
255 provide complementary information about the selection pressure acting on
256 different gene types or genomic regions. For lincRNA, pri-miRNA, and protein-
257 coding gene loci, we compared phastCons scores for exons and promoter
258 regions (2kb upstream of transcription start site) (Supplemental Data Set S2). In
259 general, for exonic sequences, lincRNAs are significantly less conserved than
260 pri-miRNAs and protein-coding genes (Figure 2B, purple, yellow, and green
261 series). However, we found that Angiosperm-, Eudicot/rosid- and
262 Brassicaceae_I_II-conserved lincRNAs were as conserved as protein-coding
263 genes. Arabidopsis-specific lincRNAs show the lowest phastCons scores.

264 In contrast to the differences in exonic sequences, the promoter scores are
265 comparable for lincRNA, pri-miRNA, and protein-coding genes (Figure 2C, purple,
266 yellow and green series). The level of purifying selection on promoter regions is
267 similar in Angiosperm- and Eudicot/rosid-conserved lincRNAs as well as in
268 protein-coding genes. We observed that conservation levels were again
269 significantly higher in the Brassicaceae_I_II-conserved lincRNA promoters than
270 protein-coding gene promoters. Taken together, these results indicate that

271 Brassicaceae_I_II-conserved lincRNAs stand out in the level of purifying
272 selection acting on these loci, both in exons and in their promoter regions,
273 suggesting that the primary sequence of the RNA and its promoter regulation is a
274 critical element for lincRNA function.

275

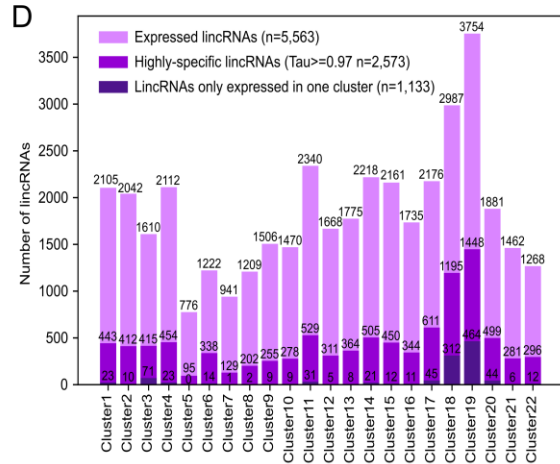
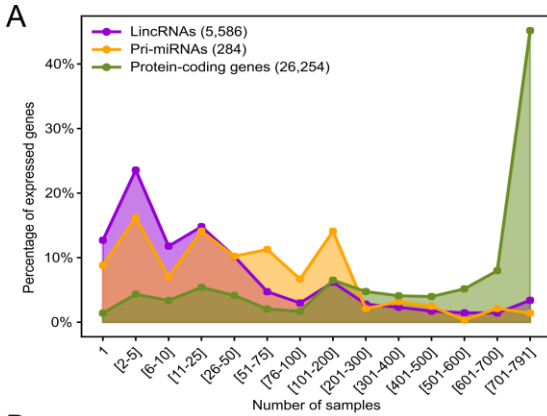
276 **Large-scale transcriptome analysis reveals highly-specific lincRNA** 277 **expression in roots**

278 Increasing evidence supports the tissue or cell-type specific role of lincRNA gene
279 functions (Liu et al., 2012; Li et al., 2016; Cheng et al., 2017; Rai et al., 2019). To
280 characterize spatiotemporal Arabidopsis gene expression patterns also
281 considering lincRNAs, we curated, processed, and integrated 791 RNA-seq
282 samples to construct a genome-wide gene expression atlas covering a wide
283 range of tissues, developmental stages, and stress conditions (Supplemental
284 Table S2 and Supplemental Data Set S3). To find a threshold of detectable
285 expression above background, we normalized all data to establish detectable
286 expression levels for protein-coding genes and lincRNAs (Ramskold et al., 2009;
287 Li et al., 2016) (see Materials and Methods, Supplemental Figure S2A). A
288 normalized transcripts per million (TPM) value ≥ 0.2 was considered to define
289 5,586 (84.6%) expressed lincRNAs and 284 (87.4%) pri-miRNAs, whereas a
290 TPM ≥ 2 was used to define 26,254 (94.9%) expressed protein-coding genes.
291 Globally, the expression breadth, defined as the number of samples in which a
292 gene is expressed, was lower for lincRNAs (median: 12/791 samples) compared
293 to pri-miRNAs (median: 35.5/791 samples) and protein-coding genes (median:
294 648/791 samples) (Figure 3A). Although the identification of lincRNAs is known to
295 be impacted by the sequencing depth (Cabili et al., 2011; Liu et al., 2012; Sun et
296 al., 2017), we did not observe a clear correlation between the number of detected
297 expressed lincRNAs and sequencing depth (Pearson's correlation coefficient,
298 $r=0.04$, $p=0.86$) (Supplemental Figure S2B). As previously reported, the
299 expression level of lincRNAs is generally lower than that of pri-miRNAs and
300 protein-coding genes (Figure 3B). The lincRNA loci part of the two "older"

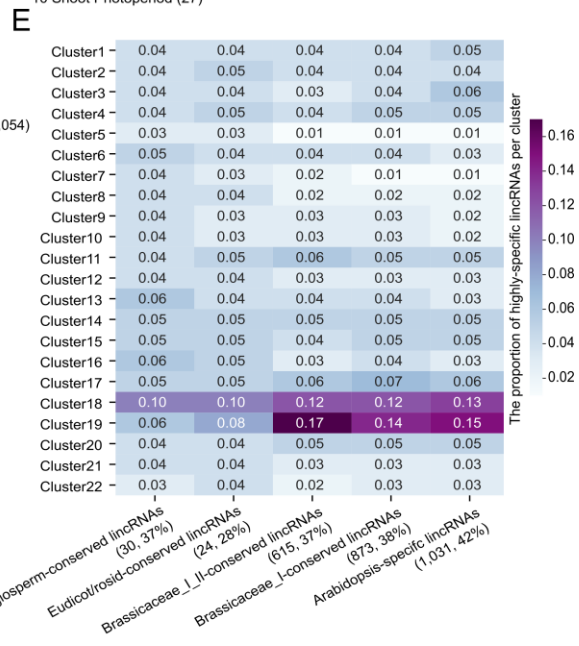
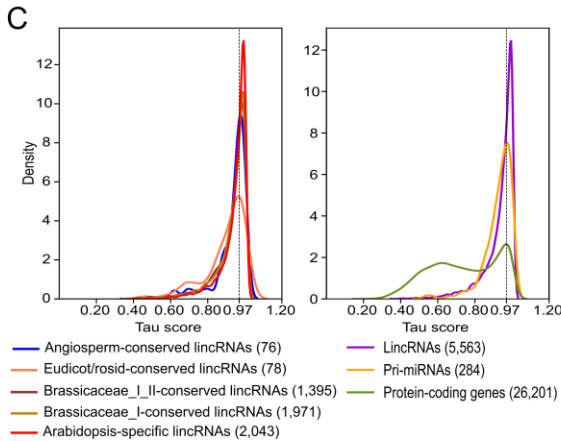
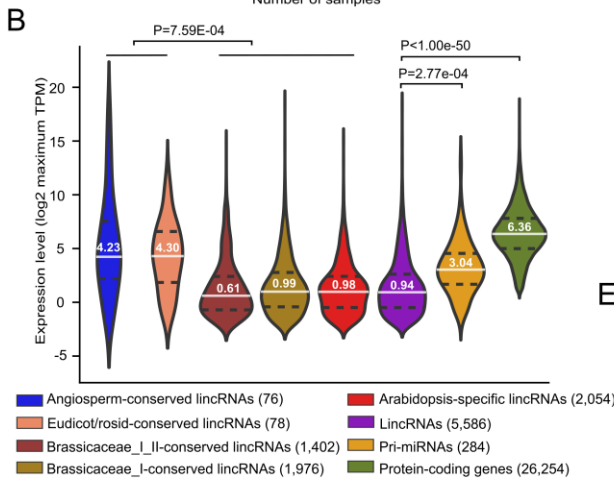
301 categories (Angiosperm and Eudicot/rosid-conserved lincRNAs) showed
302 significantly higher expression levels compared with the three “younger”
303 categories of lincRNAs (Brassicaceae_I_II-conserved, Brassicaceae_I-conserved
304 and Arabidopsis-specific lincRNA). We also observed a six-fold difference in the
305 fraction of expressed genes depending on their annotation source: lincRNAs
306 annotated in only one resource were more frequently found as undetectable
307 compared to lincRNAs annotated in two or more resources (20% versus 3%,
308 respectively).

309 To identify groups of samples with similar expression patterns, 769 RNA-Seq
310 samples were split into 22 expression clusters using meta-data curation (22
311 samples were discarded, see Materials and Methods). Globally, these clusters
312 group Arabidopsis (Col-0) samples according to the organ or tissue considered
313 and the stress, either biotic or abiotic, applied (Supplemental Data Set S3). As
314 lincRNAs are proposed to exert their role in a tissue-specific manner (Liu et al.,
315 2015), the tissue-specificity index (τ) (Yanai et al., 2005) for each lincRNA, pri-
316 miRNA and protein-coding gene was calculated to estimate expression specificity
317 between clusters (Figure 3C). As reported in other studies (Ponting and Haerty,
318 2022; Mattick et al., 2023), lincRNAs were more specifically expressed than pri-
319 miRNAs and protein-coding genes, with the median τ scores of 0.97, 0.95 and
320 0.74, respectively ($P < 0.001$, Mann–Whitney U-test). This confirms that
321 expression specificity is part of the lincRNA signature. Therefore, in the rest of
322 the study, we concentrate on lincRNAs with highly-specific expression, which we
323 defined as having a τ score greater than the 0.97 and representing forty-six
324 percent of all expressed lincRNAs (2,573/5,563) (Supplemental Data Set S4).
325 Among the highly-specific lincRNAs, the vast majority (66%) were derived from
326 one resource, with lincRNAs identified in Liu et al. 2012, an early large-scale
327 lincRNA identification studies in Arabidopsis, being most abundant (30%,
328 Supplemental Figure S3A-B). Even though we did not observe a uniform
329 distribution of these highly-specific lincRNA over the 22 clusters (Figure 3D),
330 cluster 19 (root), 18 (cell line light induction) and 3 (seedling heat) have the
331 largest number of highly-specific lincRNAs. Between 28 and 42% of the highly-

332 specific expressed lincRNAs were present in the different evolutionary categories,
 333 with the highest fraction in the Arabidopsis-specific lincRNAs (Figure 3E).
 334 Noteworthy, expression cluster 19, containing 66 samples covering both whole
 335 root and specific root cell types (Li et al., 2016), contains the highest fraction (14-
 336 17%) of highly-specific lincRNAs in the younger age categories (Figure 3E),
 337 hinting towards a role of these lincRNAs in root tissues.



- Abiotic stress**
- 1 Seedling GA and cold (38)
 - 2 Seedling Heat (32)
 - 3 Seedling Heat (21)
 - 4 Root Nutrient deficiency (26)
 - 5 Leaf Cold (5)
 - 6 Leaf Wounding (5)
 - 7 Leaf Heat (6)
 - 8 Cotyledon R:FR light (26)
 - 9 Hypocotyl R:FR light (26)
 - 10 Shoot Photoperiod (27)
- Biotic stress**
- 11 Root Bacteria (68)
 - 12 Leaf Mite infection (23)
 - 13 Leaf Bacteria (48)
 - 14 Leaf *P. syringae* (113)
 - 15 Whole plant Coronatine (83)
- Biotic/abiotic stress**
- 16 Leaf Fungus, *Pteris rapae* and drought (55)
- Tissues**
- 17 Seed (39)
 - 18 Cell line
 - 19 Root (66)
 - 20 Flower (17)
 - 21 Leaf (13)
 - 22 Silique (8)



339 **Figure 3. Expression analysis of lincRNAs.** (A) Line chart showing the
340 distribution of expression breadth for lincRNAs, pri-miRNAs and protein-coding
341 genes across all samples (B) Distribution of the maximum TPM expression levels
342 for different lincRNA age categories and gene types (lincRNAs, pri-miRNAs and
343 protein-coding genes). (C) Distribution of tissue specificity tau scores for different
344 lincRNA age categories and gene types. Tau scores range from zero to one,
345 where zero means widely expressed, and one means very specifically expressed
346 (detectable in only one cluster). The black dotted line represents a tau score of
347 0.97. (D) Bar chart reporting the number of expressed and highly-specific
348 expressed lincRNAs per expression cluster and the number of lincRNAs that are
349 only expressed in one cluster. Cluster numbers and descriptions are shown
350 below the chart, with numbers in parenthesis indicating the number of samples
351 present per cluster. (E) Heatmap showing the proportion of highly-specific
352 expressed lincRNAs in each cluster for each lincRNA age category. Cluster
353 descriptions are the same as in panel D. Numbers in parenthesis report the
354 number of genes per age category together with the fraction of lincRNAs showing
355 highly-specific expression.

356

357

358 To validate these tissue-specific expression patterns, we verified the expression
359 of previously characterized Arabidopsis lincRNAs (Supplemental Table S3). The
360 lincRNA *SVALKA*, which was identified in a cold-sensitive region of the
361 Arabidopsis genome, was maximally expressed in the cold stress-related cluster
362 5 in our study (Kindgren et al., 2018). In addition, the lincRNA *MARS*, which is
363 involved in the response to abscisic acid, seed germination and root growth
364 under osmotic stress (Roule et al., 2022b), was found to be expressed at high
365 levels in root cluster 19 as well as cluster 11 (root bacteria) and cluster 17 (seed).
366 The lincRNA *ASCO*, reported to be involved in lateral root formation and
367 response to pathogens (Bardou et al., 2014; Rigo et al., 2020), was found to be
368 widely expressed, including cluster 11, which contained samples reporting
369 bacterial flagellin stress responses. The Arabidopsis *IncCOBRA1*, also conserved

370 in field mustard (*Brassica rapa*) (Kramer et al., 2022) and playing a role in seed
371 germination, was found in our set of Brassicaceae-I_II-conserved lincRNA and
372 showed the highest expression in cluster 17 (seed) and cluster 3 (seedling heat).
373 Taken together, these findings indicate that the expression clusters offer a good
374 starting point for the context-specific characterization of known and
375 uncharacterized lincRNAs. Furthermore, young age categories, including the
376 Brassicaceae-conserved and Arabidopsis-specific lincRNAs, show a bias for
377 expression specificity in the root samples, with more than half (1,448/2,573) of
378 the highly-specific lincRNAs active in root, suggesting a potentially diverse role of
379 lincRNA networks in root growth and development.

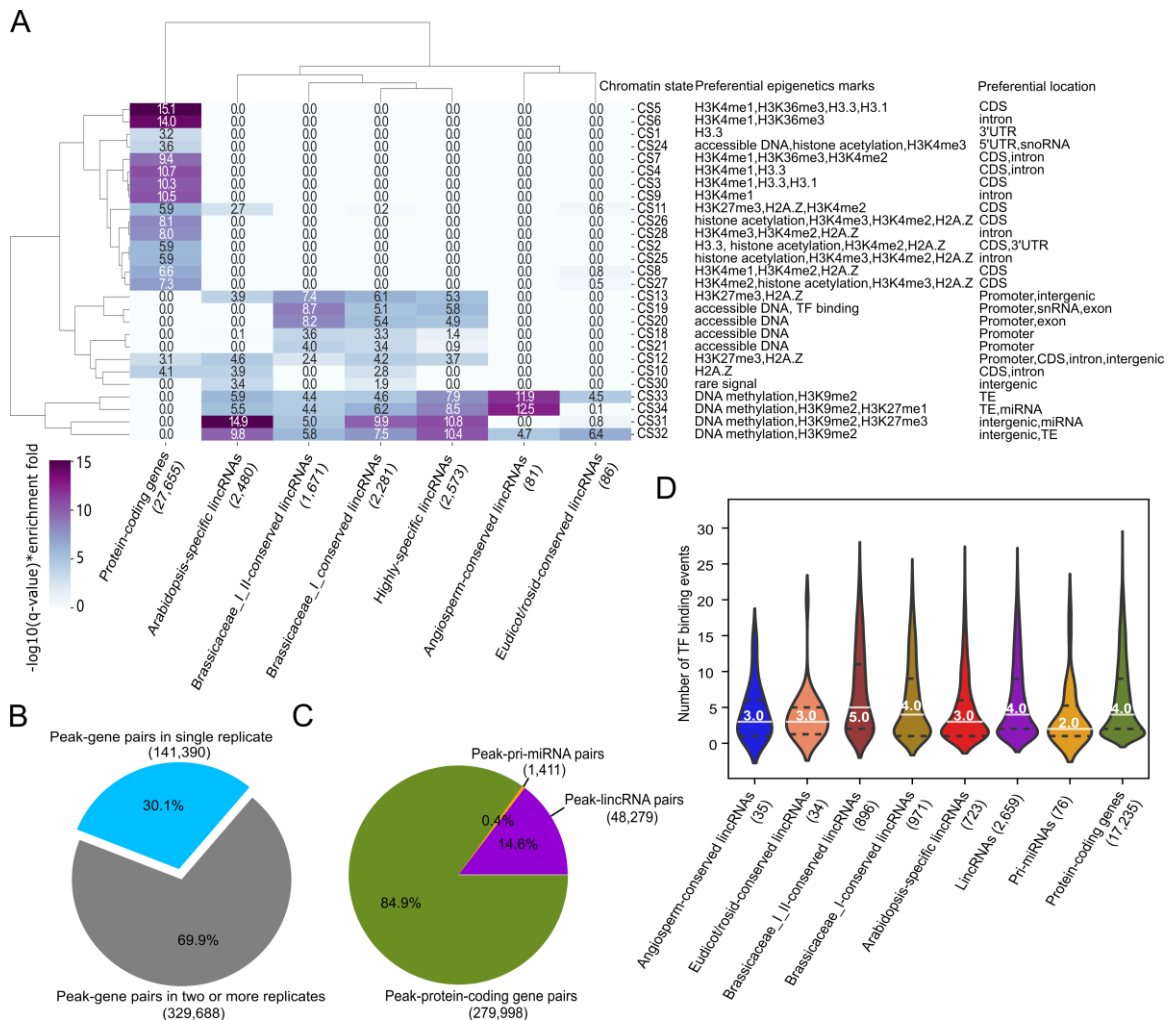
380

381 **Experimental TF-lincRNA regulatory network reveal active and complex** 382 **gene regulation of Arabidopsis lincRNA genes**

383 To integrate lincRNAs into epigenetic and transcriptional networks, we compared
384 the chromatin states inferred by (Liu et al., 2018; Hazarika et al., 2022) for the
385 different lincRNA gene sets delineated in our study. Liu and co-workers identified
386 34 different chromatin states (CS1 to CS34), consisting of different combinations
387 of epigenetic marks along the genome, which offer detailed insights in the
388 locations and functions of regulatory regions and genes (Liu et al., 2018).
389 Globally, lincRNAs show enrichment for vastly different chromatin states
390 compared to protein-coding genes (Figure 4A). Chromatin states 33-34, typically
391 associated with DNA methylation, repressive histone modifications and
392 transposable elements, were strongly overrepresented in angiosperm-conserved
393 lincRNAs, and to a lesser extent in lincRNAs showing highly-specific expression
394 patterns. Chromatin states 31-32, also associated with DNA methylation and
395 repressive histone modifications were strongly overrepresented in Arabidopsis-
396 specific lincRNAs. Interestingly, these repressive chromatin marks were less
397 enriched in the Brassicaceae-I_II-conserved lincRNAs, where chromatin states
398 13, 19 and 20 were most strongly overrepresented. State 13 refers to Polycomb
399 group mediated deposition of trimethylation of the lysine 27 of histone H3

400 (H3K27me3) while states 19-20 denote accessible chromatin and TF ChIP
401 binding. Chromatin states 19-20, but also states 31-32 and 34, were enriched in
402 highly-specific lincRNAs. Although we found no evidence of a correlation
403 between lincRNA age categories and chromatin states, our results revealed that
404 a significant number of Brassicaceae_I_II-conserved lincRNAs have epigenetic
405 signatures associated with Polycomb regulation and TF binding in accessible
406 chromatin.

407 To refine which chromatin states are more associated with active versus
408 repressed lincRNAs, we defined lincRNA gene sets with different expression
409 patterns and compared their chromatin state enrichments. For lincRNAs showing
410 active expression in root, various chromatin states (CS 10-13, CS 20) linked with
411 promoter, coding sequences, and introns, containing epigenomic marks for
412 H3K27me3, H2A.Z, and accessible DNA, were found to be strongly enriched
413 compared to non-expressed lincRNAs (Supplemental Figure S4A). Reversely,
414 non-expressed and non-root-expressed lincRNAs showed strong enrichment for
415 H3K9m2, DNA methylation, and H3K27me1 (CS 32-34), marks frequently found
416 in intergenic regions and transposable elements, which agrees with their role in
417 silencing and controlling DNA methylation.



418

419

Figure 4. Chromatin state and TF ChIP-Seq peak annotation for different gene types (A) Dendrogram showing the enrichment for different lincRNA gene sets (x-axis) towards different chromatin states (CS) (y-axis). The values report the product of $-\log_{10}(\text{q-value})$ and the enrichment fold. Only significant enrichment values are reported ($\text{q-value} < 0.05$). **(B)** The proportion of peak-gene pairs present in single replicate (blue) and two or more ChIP-Seq replicates (grey). **(C)** The percentage of three gene types assigned to peaks in two or more ChIP-Seq replicates. **(D)** Distributions of the number of TF binding events for lincRNA evolutionary age categories and gene types (lincRNAs, pri-miRNAs and protein-coding genes).

429

430

431 Based on the specific expression profiles for different lincRNA genes, the
432 chromatin state information, as well as the high levels of promoter conservation,
433 we next integrated TF chromatin immunoprecipitation (ChIP) data to further
434 identify the regulators controlling lincRNAs. Before we used TF ChIP data to
435 characterize the organization, complexity, and evolution of TF binding for protein-
436 coding genes (Heyndrickx et al., 2014), hence we here re-processed publicly
437 available ChIP-Seq to identify TF binding events potentially controlling lincRNA
438 gene expression. A total of 114 TF ChIP-Seq datasets, covering 45 TFs with at
439 least two biological replicates, were reprocessed (Supplemental Table S4).
440 Starting from our genome annotation containing protein-coding genes, pri-
441 miRNAs and the high-confidence lincRNAs, ChIP-Seq peaks were assigned to
442 the closest genes. A gene was defined as a potential target gene for a profiled TF
443 if it was the closest to at least one ChIP-Seq peak of that TF (see Materials and
444 Methods). Based on all 471,078 TF peak-target gene pairs identified from the
445 114 ChIP-Seq datasets, 329,688 (70%) of the peak-gene pairs were confirmed
446 by two or more replicates and were kept to construct a robust TF- target gene
447 regulatory network (Figure 4B). While most TF peaks were associated with
448 protein-coding genes, we identified 48,279 (14.6%) peaks that were associated
449 with lincRNAs (Figure 4C). Given the strong localization bias for TF binding in
450 Arabidopsis (Heyndrickx et al., 2014; Yu et al., 2016), we only considered
451 expressed target genes localized relative to the peak midpoint within a 2kb
452 window (Supplemental Figure S4B), retaining 93.0% of the binding events
453 (15,686 TF-lincRNA interactions, see Supplemental Data Set S5). Globally, 2,659
454 expressed lincRNAs had one or more TF binding events. Twenty-two out of the
455 45 TFs, belonging to the bHLH, HD-Zip, bZIP, C2H2, GRAS, MYB, NAC, NF-YB
456 and NF-YC TF families, were each associated with at least 200 lincRNAs
457 (Supplemental Table S5).

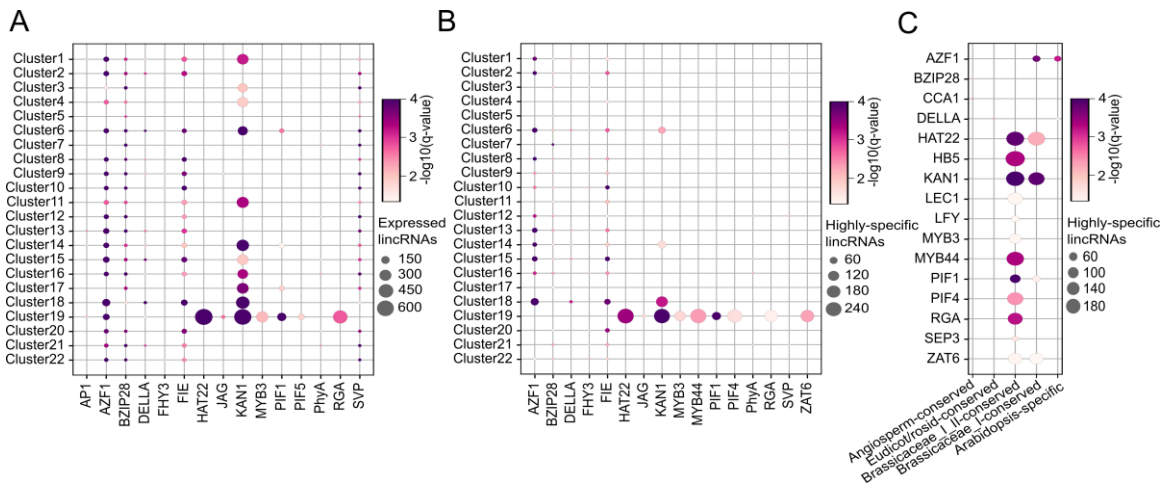
458 (Haudry et al., 2013) identified over 90,000 conserved non-coding sequences
459 (CNS) in Arabidopsis that show evidence of transcriptional and post-
460 transcriptional regulation. Comparing these CNS with the ChIP-Seq peaks of
461 target genes revealed that most of the TF binding events close to lincRNAs

462 (78.1%) and protein-coding genes (73.7%) overlapped with a CNS
463 (Supplemental Figure S4C). This fraction was much lower for pri-miRNAs
464 (57.0%). The highest fraction of ChIP-Seq peaks containing a CNS was detected
465 for Brassicaceae_I_II-conserved lincRNAs loci (91.8%), indicating that these TF
466 binding events are evolutionary constrained and potentially functional.
467 Considering the different gene types, the median number of TF binding events
468 per locus was higher for protein-coding genes and lincRNAs (median of 4 TFs)
469 compared to pri-miRNAs, suggesting these genes are differently controlled
470 (Figure 4D). Brassicaceae-conserved lincRNAs have more TF binding events
471 than lincRNAs in any other age categories (Figure 4D). Furthermore, we
472 observed a positive correlation between TF binding frequency and the
473 conservation of lincRNA gene body ($r=0.27$, $P=8.11e-92$) (Supplemental Figure
474 S4D). More precisely, a large fraction of lincRNAs without any TF binding event
475 also has very low phastCons scores, close to zero, indicating that these genes
476 are not under purifying selection. We observed a negligible correlation between
477 the number of TF binding events and tissue specificity ($r=0.02$, $p=1.69e-01$).
478 Altogether, lincRNAs experiencing stronger levels of purifying selection are
479 bound, and potentially regulated, by more TFs, independent of their tissue
480 specificity (Supplemental Figure S4E). Taken together, we generated a multi-
481 level genome-wide characterization covering chromatin state information,
482 promoter conservation, ChIP-based TF binding and CNSs, for all detectable
483 lincRNA across >700 expression samples, permitting to rapidly define the
484 biological context and relevance of lincRNAs in Arabidopsis regulatory networks.
485

486 **MYB44, PIF4 and KAN1 regulate Arabidopsis lincRNAs in different root cell** 487 **types**

488 To identify TF-lincRNA regulatory interactions active in specific cellular contexts,
489 we used the previously defined expression clusters to combine TFs regulation
490 and lincRNA expression in specific organs, tissues, or stress conditions.
491 Considering all expressed lincRNAs and all TF peak-based regulatory

492 interactions described above, we tested if specific expression clusters are
 493 overrepresented for lincRNAs controlled by specific TFs (Y-axis and X-axis in
 494 Figure 5A-B, respectively). We observed a significant overrepresentation for
 495 KANADI1 (KAN1) binding to lincRNA loci in 11 clusters, of which clusters 19
 496 (root), 18 (cell line light induction), 14 (leaf *P. syringae*) and 6 (leaf wounding)
 497 showed the most significant overlaps (Figure 5A). Comparing the different
 498 clusters revealed that root cluster 19 contained numerous enriched TFs,
 499 including Arabidopsis Zinc-Finger protein 1 (AZF1), JAGGED (JAG),
 500 Phytochrome-Interacting Factor 5 (PIF5), Repressor of GA (RGA), PIF1, MYB3,
 501 KAN1 and HAT22. The observed patterns of TF-binding in different expression
 502 clusters were unique for lincRNAs and highly dissimilar compared to TF binding
 503 enrichment for protein-coding genes (Supplemental Figure S5A).
 504



505
 506 **Figure 5. Overview of TF-lincRNA regulatory interactions in different**
 507 **expression clusters and age categories.** The dot sizes represent the number
 508 of the lincRNAs while the color represents the statistical significance. Cluster
 509 descriptions are the same as in Figure 3D. **(A)** Bubble chart showing the
 510 enrichment of TF binding for expressed lincRNAs in different expression clusters.
 511 TFs lacking significant enrichment in any of the 22 clusters are not shown. **(B)**
 512 Bubble chart showing the enrichment of TF binding for highly-specific lincRNAs in
 513 different expression clusters. TFs lacking significant enrichment in any of the 22

514 clusters are not shown. **(C)** Bubble chart showing the enrichment of TF binding
515 for expressed lincRNAs in different age categories.

516

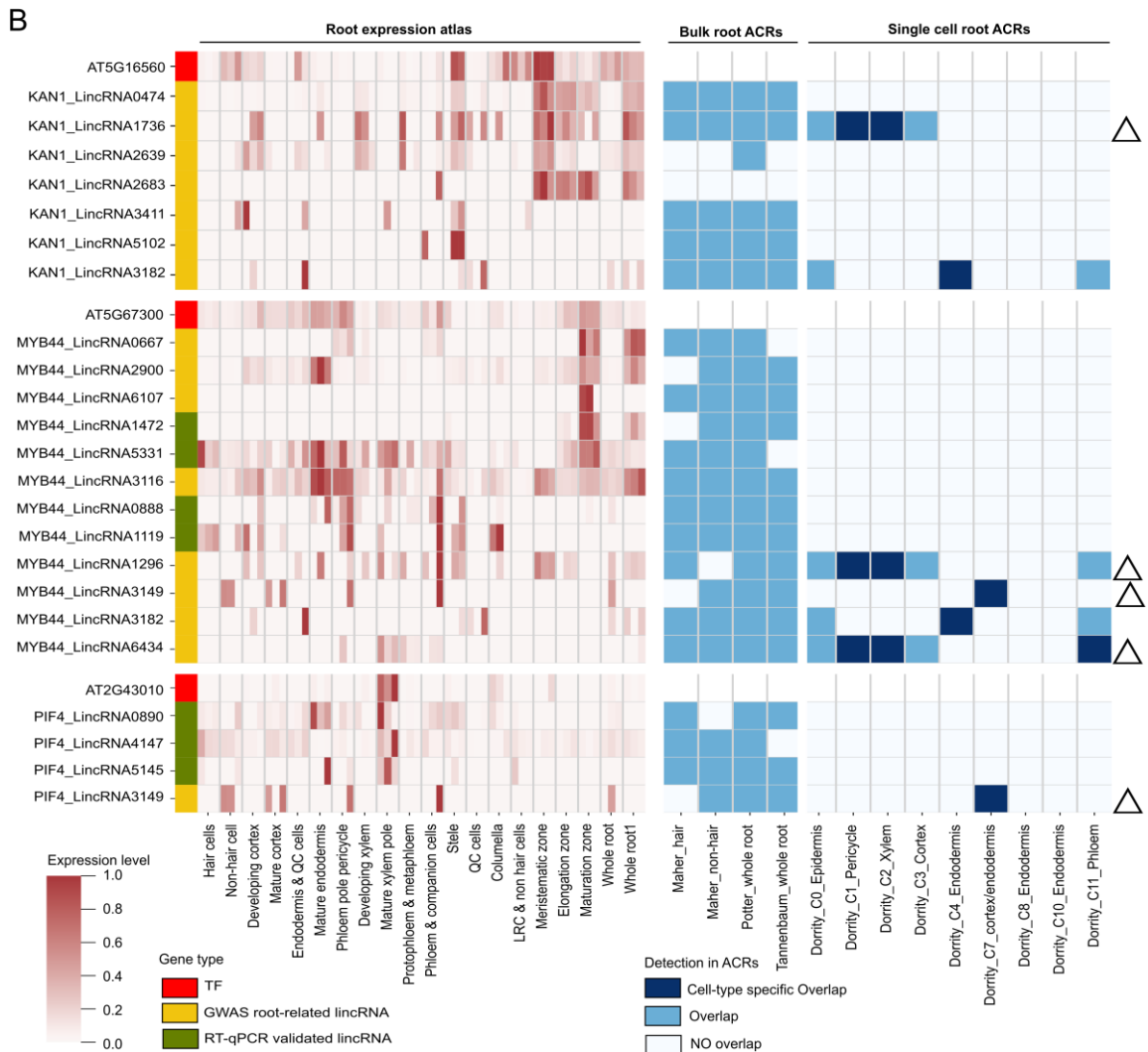
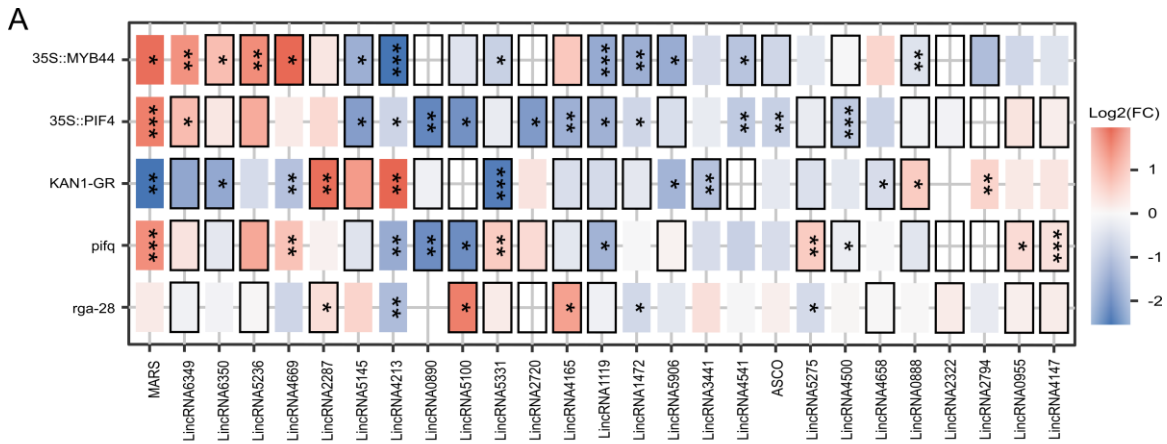
517

518 Focusing on the set of 2,573 lincRNAs showing highly-specific expression, most
519 of the TF enrichments for root cluster 19 remained (Figure 5B). PIF1, PIF4, KAN1,
520 HAT22, MYB3, MYB44, ZAT6 and RGA showed the largest overlap and all these
521 TFs, apart from MYB3 and PIF1, contained >160 lincRNA target genes (Figure
522 5B and Supplemental Table S6). We confirmed that these eight TFs were all
523 expressed in one or more samples of the root expression cluster 19
524 (Supplemental Figure S5B) and earlier studies have reported that these TFs,
525 apart from HAT22, are involved in root development in Arabidopsis (Hawker and
526 Bowman, 2004; Devaiah et al., 2007; Moubayidin et al., 2016; Tominaga-Wada
527 and Wada, 2016; Zhao et al., 2016; Li et al., 2022). Comparing TF binding for the
528 different age categories revealed that the Brassicaceae_I_II-conserved lincRNAs
529 were most strongly enriched for these TFs (Figure 5C).

530 To experimentally validate the ChIP-based regulatory interactions for the
531 identified TFs, we used reverse transcription quantitative PCR (RT-qPCR)
532 analysis in roots of lines affected in TF expression such as overexpression or T-
533 DNA inactivation (called TF perturbation lines). We used an inducible line for
534 KAN1 (*KAN1-GR*), overexpression lines for MYB4 and PIF4 (*35S::MYB4*,
535 *35S::PIF4*) or quadruple mutant of the *pif1*, *pif2*, *pif3*, and *pif4* (*pifq*), and a
536 knockout mutant for RGA (*rga28* T-DNA line). We then selected 27 potentially
537 regulated lincRNAs showing high expression levels in the root cluster 19 and that
538 were targeted by several of the selected TFs. For example, *LincRNA5331* and
539 *LincRNA1119* were predicted to be regulated by the four TFs. We could detect
540 deregulation for *LincRNA5331* and *LincRNA1119* in *35S::MYB44* and *pifq* roots,
541 whereas the former was also deregulated in *KAN1-GR* and the latter in *35S::PIF4*
542 (Figure 6A, Supplemental Figure S6). Overall, out of the 74 inferred regulatory
543 interactions investigated, 36 were confirmed, meaning that for the tested TFs, a
544 significant deregulation of the lincRNA was found in comparison to the control

545 (“Has a peak and is DE”, in Table 1, Supplemental Table S7). For 23 out of the
546 27 tested lincRNAs, we confirmed one or more regulatory interactions (Figure
547 6A). The precision (i.e. the proportion of regulation prediction that were confirmed
548 by RT-qPCR experiment), varied between 27-65% and the recall (i.e. the
549 proportion of regulation seen by RT-qPCR that were correctly predicted by ChIP-
550 seq data) varied between 60% and 100%, while the average accuracy (i.e. the
551 proportion of correct predictions, regulation and absence of regulation, among all
552 genes examined) was 59%. For eight interactions the TF peak annotation to the
553 lincRNA was unclear (“Has a putative peak and is DE” in Table 1), meaning
554 these deregulated lincRNAs might also resemble confirmed interactions. Lastly,
555 while the 10 interactions where we observed deregulation in the absence of a
556 peak could indicate false predictions, they might also represent cases of indirect
557 regulation controlled by the perturbed TF, influencing the expression of the
558 profiled lincRNAs.

559 We further validated our TF-gene regulatory interactions using *in vitro* DNA
560 affinity purification sequencing (DAP-seq) data (O'Malley et al., 2016). DAP-Seq
561 peaks, available for 9 TFs included in our analysis, were retrieved from the Plant
562 Cistrome Database and were assigned to the closest gene (within a 2kb
563 distance). Overall, 40% (1,735/4,308) of the ChIP-based TF-lincRNA interactions
564 were confirmed by DAP-seq. For MYB44, one of the TFs for which we
565 experimentally validated regulatory interactions and for which DAP-Seq data is
566 available, 5/11 (45%) of the RT-qPCR confirmed lincRNAs were confirmed (two
567 examples are shown in Supplemental Figure S7). Given the technical failures
568 associated with DAP-Seq for some TFs, these confirmation rates are in
569 agreement with the overlaps reported for protein-coding genes for DAP- and
570 ChIP-Seq (36-81%, (O'Malley et al., 2016)), corroborating the quality of the
571 reported TF–lincRNA interactions.



572

573

574

575

576

Figure 6. Experimental validation and characterization of TF-lincRNA regulatory interactions. (A) Heatmap showing log₂ fold change (FC) of lincRNA relative expression levels in transcription factor (TF) overexpressing lines (35S::MYB44, 35S::PIF4 and KAN1-GR) or TF knockout lines (*rga28* and *pifq*

577 (*pif1/pif2/pif3/pif4* quadruple mutants)) vs. control wild-type lines in 14-day old
578 roots. Expression values were determined by RT-qPCR. Asterisks indicate
579 statistically significant differences (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$) in an
580 unpaired two-tailed Student's t-Test ($n = 3$). Solid boxes indicate TF ChIP peaks
581 <2kb from the lincRNA gene. Dashed boxes indicate TF ChIP peaks were
582 identified only in one ChIP-Seq replicate or the lincRNA is not the closest gene to
583 the TF peak. **(B)** Heatmap showing cell type-specific expression of lincRNAs
584 consistent with the expression of the regulatory TF, together with root ACR
585 information. Yellow and green report the GWAS root-related and RT-qPCR
586 experimentally validated lincRNA genes, respectively. The color scale represents
587 the expression levels of lincRNAs in root cells. Triangles indicate cases where
588 lincRNA expression and ACRs confirm regulatory interaction in the same cell
589 type.

590

591 **Functional and cell-type specific annotation of root lincRNAs**

592 Based on the experimental validation results confirming that several of the root-
593 specific lincRNAs are controlled by the TFs inferred using the ChIP-based peaks,
594 we integrated genotype-phenotype relationships from the AraGWAS catalog, to
595 verify if root-related phenotypes have been reported for regions containing
596 lincRNAs in genome-wide association studies (GWAS). After processing the
597 significant associations from all GWAS studies present in the catalog and only
598 retaining associations overlapping with lincRNA gene bodies or 2kb promoter
599 regions (see Materials and Methods), we identified 2,615 single nucleotide
600 polymorphisms (SNPs) overlapping with 1,039 lincRNAs, covering 142 different
601 studies. While 58.4% of the lincRNAs had only significant associations via SNPs
602 in the promoter region, 41.5% had associations via SNPs in the gene body (207
603 and 225 lincRNAs with significant associations only in the gene body and in both
604 gene body and promoter, respectively). After parsing and summarizing
605 phenotype information from the available trait ontology annotations, we could link
606 20 lincRNAs to abiotic stress-related traits, 124 lincRNAs to flower-related traits,

607 29 lincRNAs to leaf-related traits, 339 lincRNAs to root-related traits, and 25
608 lincRNAs to seed-related traits (see Supplemental Data Set S6). The number of
609 root-related GWAS annotations was much larger for lincRNAs targeted by TFs
610 showing significant enrichment for binding in root cluster 19, compared to TFs
611 lacking this enrichment (Supplemental Figure S8A), confirming the functional
612 relevance of these lincRNAs in roots. For the 339 lincRNAs with root-related
613 traits, a significant genetic association exists between a lincRNA-associated
614 sequence polymorphism and a root-related phenotype that was quantified.
615 Examples include lincRNAs affecting root mass density (48 genes) and lateral
616 root length or number (15 genes) (Supplemental Figure S8B). The significant
617 SNPs related to root traits were found in 32.1% and 67.9% of lincRNA gene
618 bodies and promoter regions, respectively (Supplemental Figure S8C). Moreover,
619 we identified several lincRNAs with specific root phenotypes regulated by MYB44
620 (Supplemental Figure S8 D-G). Overall, the reprocessed GWAS data indicate
621 that hundreds of regions overlapping with lincRNAs loci are significantly
622 associated with different plant traits and can be used to prioritize lincRNAs likely
623 involved in specific biological processes.

624 To further investigate the biological relevance of the identified regulatory network,
625 we focused on KAN1, MYB44, and PIF4, as these regulators had several root-
626 expressed lincRNA target genes that showed differential expression in TF
627 perturbation lines (Table 1). Starting with those lincRNAs that were validated by
628 RT-qPCR experiments or had root-related traits in the GWAS catalog, we first
629 screened for lincRNAs expressed in root cluster 19 and in at least two replicates
630 of a specific root cell type. A total of 21/42, 25/45, 23/45 lincRNAs fulfilling this
631 selection criteria were regulated by KAN1, MYB44 and PIF4, respectively. Next,
632 we verified the cell-type specific lincRNA expression agreed with the expression
633 of the regulatory TF (Figure 6B). For KAN1, which showed the highest
634 expression in the meristematic zone and stele, we identified six lincRNAs with
635 root-related traits showing similar cell-type specific expression. For MYB44, we
636 identified eight confirmed lincRNAs targets (four GWAS root-related and four
637 experimentally validated genes) with high expression levels in the maturation

638 zone and phloem pole pericycle. The high expression of PIF4 in the mature
639 xylem pole agrees with three experimentally validated lincRNAs. All GWAS root-
640 related (66) or experimentally validated (23) lincRNAs regulated by KAN1,
641 MYB44 or PIF4 can be found in Supplemental Table S8.

642 To confirm the root and cell-type specificity of the TF-lincRNA regulatory
643 interactions, we integrated publicly available chromatin accessibility datasets of
644 Arabidopsis roots based on Assay for Transposase-Accessible Chromatin using
645 sequencing (ATAC-seq), covering three bulk datasets (Maher et al., 2018; Potter
646 et al., 2018; Tannenbaum et al., 2018) and one single-cell dataset (Dorrity et al.,
647 2021). Assessing the tissue specificity of our TF-lincRNA regulatory network
648 revealed that 93% of the TF peaks associated with lincRNAs were detected in
649 root accessible chromatin regions (ACRs), of which 21% were specifically
650 detected in cell-type specific root ACRs. Randomly shuffling the TF peaks
651 associated with lincRNAs showed that the observed overlap of TF-lincRNA
652 regulatory interactions with bulk and single-cell root ACRs was 10-fold and 36-
653 fold higher than expected by chance, respectively, confirming the high specificity
654 of the inferred regulatory interactions. Of the 69 regulatory interactions covering
655 GWAS root-associated or experimentally validated lincRNAs regulated by KAN1,
656 MYB44, or PIF4, 67 were detected in the bulk root ACRs, of which 7 were
657 detected in root cell-type specific ACRs (Figure 6B, Supplemental Table S8). For
658 several lincRNAs the cell-type specificity identified using the gene expression
659 profiles were confirmed by the single-cell ATAC data. For example,
660 *LincRNA1736*, regulated by KAN1, and *LincRNA1296* and *LincRNA6434*,
661 regulated by MYB44, were expressed in stele (xylem/phloem) and confirmed by
662 xylem-specific ACRs. *LincRNA3149*, regulated by MYB44 and PIF4, was
663 expressed in cortex, which was confirmed by cortex-specific ACRs.

664 This detailed regulatory annotation further supports that KAN1, MYB44 and PIF4
665 are controlling lincRNA genes showing highly specific expression in different root
666 tissues and cell types. Additionally, the large overlap with genetic associations
667 from different GWAS studies hints to a role for many of these TF-controlled
668 lincRNA loci in root growth and development.

669

670

671 **DISCUSSION**

672 **Comprehensive annotation of Arabidopsis lincRNAs using transcriptomics** 673 **and evolutionary genomics**

674 In contrast to protein-coding genes, the characterization of lincRNAs is more
675 challenging as we lack highly curated annotations and extensive experimental
676 observations. Furthermore, the low levels of sequence conservation for the
677 majority of lincRNA loci makes it difficult to translate biological knowledge learned
678 in one species to another. Apart from co-expression network analysis, reporting
679 putative (in) direct associations between lincRNAs and other genes, information
680 about TF regulation of lincRNAs is scarce. Embedding lincRNAs in biological
681 networks has great potential to define lincRNAs linked to specific cellular or
682 morphological phenotypes.

683 Through the integration of eight lincRNA annotation resources, as well as
684 mapping various conservation, chromatin, and expression features, we presented
685 a global view on gene regulation for 5,586 expressed Arabidopsis lincRNAs. We
686 strongly focused on using replicated samples when processing high-throughput
687 datasets to obtain high-confidence gene information (Ponting and Haerty, 2022).
688 Combined with comparative genome analysis yielding information about age
689 categories and selection acting on gene bodies and promoter regions, we found
690 that different subsets of lincRNAs have distinct molecular properties. While the
691 high tissue-specificity and low levels of primary sequence conservation
692 corroborate previous findings about lincRNAs (Necsulea et al., 2014; Ma et al.,
693 2019; Palos et al., 2022), the analysis of lincRNA expression using a genome-
694 wide gene expression atlas covering 769 samples revealed that lincRNA
695 expression is widespread in different organs and not restricted to stress
696 conditions, which is in agreement with previous studies (Jha et al., 2020; Corona-
697 Gomez et al., 2022). In a recent study, Corona-Gomez and colleagues
698 reconstructed a co-expression network to annotate lincRNAs with associated

699 protein-coding genes. They identified several modules associated with root
700 development or root-related stress functional annotation (Corona-Gomez et al.,
701 2022). These results are consistent with the high representation of lincRNAs
702 exhibiting highly-specific expression in the root expression cluster (unique set of
703 1448 lincRNAs), which is the highest number among all expression clusters we
704 studied. When intersecting the different age categories with the expression
705 clusters, Brassicaceae_I_II-conserved lincRNAs covered a large fraction of these
706 root-expressed lincRNAs. While recent studies also identified a large number of
707 context-specific lincRNAs expressed in root tip or meristem (Corona-Gomez et al.,
708 2022; Palos et al., 2022), our age category analysis revealed that many lincRNAs
709 that originated in the common ancestor of Brassicaceae lineages I and II, showed
710 specific expression in various root cell types.

711 We observed a clear trend of increasing levels of lincRNA gene expression when
712 going from young to old age categories. However, quantifying selection levels
713 acting on different loci revealed that the oldest age categories, showing the most
714 wide-spread and highest expression, are not experiencing the highest levels of
715 purifying selection. While older categories like angiosperm-conserved and
716 Eudicot/rosid-conserved lincRNAs had similar median phastCons scores as
717 protein-coding genes, Brassicaceae_I_II-conserved lincRNAs showed the higher
718 levels of purifying selection, both in their exon and promoter. In animals, a
719 substantial increase of dynamically expressed genes and higher levels of
720 purifying selection has been reported for older age categories (Necsulea et al.,
721 2014; Sarropoulos et al., 2019). Furthermore, massively parallel reporter assays
722 surveying thousands of human promoters revealed that tissue-specific lincRNAs
723 had fewer TF motifs compared to ubiquitously expressed genes (Mattioli et al.,
724 2019). In contrast to Arabidopsis protein-coding genes, this pattern was not found
725 for lincRNAs in our analysis (correlation between tau score and TF binding
726 frequency = 0.02). These results suggest, based on the genome-wide TF binding
727 data available in Arabidopsis, that the complexity in TF control of lincRNAs is
728 different between plants and animals and that the observed pattern of highly-
729 specific expression and purifying selection for Brassicaceae_I_II-conserved

730 lincRNAs deviates from the global trends observed in animals. Therefore, these
731 unique properties suggest that these lincRNAs, which only emerged 42 million
732 years ago, are better integrated in plant networks when compared to young
733 lincRNAs in animals.

734

735 **Integrative regulatory annotation of Arabidopsis lincRNAs**

736 The regulatory annotation using genome-wide chromatin state information
737 revealed that protein-coding genes and lincRNAs have distinct chromatin
738 signatures. The enriched chromatin states for lincRNAs were largely variable
739 between, and sometimes within, the different age categories. While states
740 associated with DNA methylation and repressive histone modifications were most
741 strongly overrepresented in the youngest (Arabidopsis-specific) and oldest
742 (angiosperm-conserved) lincRNA age categories, states denoting polycomb
743 group mediated deposition of H3K27me3 and accessible chromatin were most
744 enriched for Brassicaceae_I_II-conserved lincRNAs. H3K27me3 is a repressive
745 covalent histone modification resulting from the activity of Polycomb repressive
746 complexes. It was recently shown that a reduction in H3K27me3 levels leads to a
747 decrease in the interactions within Polycomb-associated repressive domains,
748 resulting in a global reconfiguration of chromatin architecture and transcriptional
749 reprogramming during plant development (Huang et al., 2021). Chromatin
750 accessibility is a hallmark of regulatory DNA as it allows sequence-specific
751 binding of TFs, key components of transcriptional regulatory networks (Schmitz
752 et al., 2022). The association of these chromatin states with specific sets of
753 lincRNAs strongly indicates active transcriptional regulation.

754 To identify context-specific TF regulation potentially driving the highly-specific
755 expression observed for many lincRNAs, we reprocessed, filtered and annotated
756 114 ChIP-Seq experiments covering 45 TFs, yielding a TF-lincRNA gene
757 regulatory network containing 2,659 lincRNAs and 15,686 interactions. To assess
758 the potential functionality of these inferred regulatory interactions, we overlapped
759 CNSs identified using nine Brassicaceae genomes, which confirmed that TF

760 peaks close to lincRNAs show similarly high levels of sequence constraint
761 compared to protein-coding genes (74-78%). Furthermore, TF binding events
762 close to Brassicaceae_I_II-conserved lincRNAs showed the highest levels of
763 CNS conservation (92%), which agrees with the very high phastCons promoter
764 scores we observed for this age category. While Palos and co-workers reported
765 that CNSs significantly correlated with gene bodies of Brassicaceae-conserved
766 lincRNAs (Palos et al., 2022), our results revealed that also lincRNA promoters
767 and TF binding sites are strongly conserved for Brassicaceae_I-II-conserved
768 lincRNAs. Ultraconserved CNSs, frequently associating with TF binding sites for
769 key plant regulators controlling essential biological processes, have been
770 identified for thousands of protein-coding genes (Van de Velde et al., 2016). Our
771 results suggest that such deep conservation of cis-regulatory elements is
772 extremely rare for lincRNAs, as only a small number of lincRNA genes show
773 deep evolutionary conservation. Such deeply conserved CNSs for protein-coding
774 genes frequently occur in divergent gene pairs, where they form mini-regulons
775 representing conserved transcriptional units of co-regulated and co-expressed
776 neighboring genes (Van de Velde et al., 2016). It is currently unclear if such
777 conserved transcriptional regulons also exist for lincRNA loci and could explain
778 the observed patterns of positionally-conserved but sequence-diverged lincRNAs
779 (Mohammadin et al., 2015).

780 The integrated TF binding information showed that promoters of lincRNAs differ
781 strongly from those of protein-coding genes, but also revealed high levels of
782 heterogeneity among the different age categories. In animals, promoters of
783 protein-coding genes contain more TF binding sites than those of lincRNAs,
784 suggesting a stronger and more complex transcriptional regulation of the former
785 (Necsulea et al., 2014; Sarropoulos et al., 2019). When comparing the number of
786 TF binding events for the different gene types in Arabidopsis, we observed no
787 difference in TF binding frequency for protein-coding genes and lincRNAs,
788 indicating that lincRNAs are also regulated in a complex manner in plants. While
789 no correlation between lincRNA expression tissue-specificity and TF binding
790 frequency was found, a positive correlation between TF binding frequency and

791 the level of purifying selection was observed. Again, Brassicaceae_I_II-
792 conserved lincRNAs stood out having the highest number of binding TFs,
793 suggesting that neither the age nor the expression of a lincRNA, but its
794 importance for plant fitness, is a major factor in determining its regulatory
795 complexity. While our results confirm that broadly expressed protein-coding
796 genes, showing high expression breadth, are positively correlated with the
797 number of regulating TFs (Heyndrickx et al., 2014), we did not observe this global
798 trend for lincRNAs, indicating that the regulatory properties of TF control for plant
799 protein-coding genes and lincRNAs are different. Although the number of
800 Arabidopsis TFs profiled using CHIP-Seq may be considered as limited, future
801 research will have to address whether the complexity of TF control varies for
802 lincRNAs, as well as protein-coding genes, active in different organs, tissues, or
803 stress conditions.

804

805 **A TF-lincRNA gene regulatory network identifies KAN1, MYB44 and PIF4 as** 806 **regulators controlling root lincRNAs**

807 Through integration of our spatiotemporal expression clusters and the TF-
808 lincRNA gene regulatory network, we identified eight TF regulators showing a
809 significant enrichment for TF binding close to lincRNAs specifically expressed in
810 roots. While the overlap between TF CHIP-Seq peaks and CNSs gave an indirect
811 indication of the potential functionality of TF binding sites close to hundreds of
812 lincRNA loci, we experimentally validated a set of inferred regulatory interactions,
813 focusing on 27 root-expressed lincRNAs and 5 TF perturbation lines. The number
814 of confirmed regulatory interactions as well as the positive prediction values
815 found for the tested TFs and lincRNAs here (27-65%) are 3 to 4 times higher
816 than the fraction of TF-bound protein-coding genes also showing deregulation
817 reported for a set of TF regulators involved in flowering (7-22%) (O'Maoileidigh et
818 al., 2014). Compared to the discovery rates obtained for large-scale phenotypic
819 screens of insertional lines (1.3%) (Ransbotyn et al., 2015), our discovery rates
820 for deregulation are 20-50 fold higher. Globally, for 85% of the profiled lincRNAs

821 one or more confirmed regulatory interactions were found. These findings
822 indicate that the ChIP peak-based inference of TF regulation is a promising
823 approach to characterize TF regulation of lincRNAs. Additional deregulated
824 lincRNAs lacking a TF peak were also identified, which might be due to an
825 indirect effect caused by crosstalk between different regulators in the Arabidopsis
826 root as well as the type of mutations chosen in the perturbed TF lines (e.g.
827 partners lacking in overexpressing lines, compensatory effects in gene families)
828 (Heyndrickx et al., 2014).

829 While detailed functional characterizations of lincRNA genes are scarce, the
830 integration of GWAS information allowed us to identify genetic associations for
831 1,039 lincRNAs covering various traits. While this number, corresponding to a
832 frequency of 8%, is slightly lower compared to that of protein-coding genes
833 associated with a specific trait in the AraGWAS catalog ($3,030/27,655 = 11\%$), it
834 does confirm the great potential of this largely untapped resource to biologically
835 characterize lincRNAs potentially controlling different plant traits. As shown for
836 KAN1, MYB44 and PIF4, multiple groups of co-expressed lincRNAs were
837 identified bound by one of these TFs. Most of these groups showed strong cell-
838 type specific expression and contained lincRNAs that were annotated with root-
839 related traits. While most regulatory interactions were confirmed by TF peaks
840 overlapping with root bulk ACRs, for several lincRNAs cell-type specificity was
841 confirmed based on single-cell ATAC data, despite the high sparsity associated
842 with this data type.

843 While co-expression networks and modules containing lincRNA genes cannot
844 differentiate between direct and indirect regulatory interactions and lack
845 functional information about individual lincRNAs (Corona-Gomez et al., 2022;
846 Palos et al., 2022), our complementary approach relying on TF regulation and
847 GWAS information overcomes some of these shortcomings. Taken together, the
848 integration of different gene annotations combined with information about
849 evolutionary conservation, selection, expression, TF regulation and GWAS data
850 yielded insights on the biological relevance of hundreds of Arabidopsis lincRNAs

851 and offers a promising strategy to identify lincRNAs involved in different aspects
852 of plant biology.

853

854

855 **MATERIALS AND METHODS**

856 **Prediction of lincRNAs from in-house dataset**

857 Paired-end RNA-seq datasets with high sequencing depth conducted in previous
858 projects in the group at the Institute of Plant Sciences Paris-Saclay were used to
859 predict additional lincRNAs. All data came from experiments carried out in the
860 *Arabidopsis (Arabidopsis thaliana)* Col-0 ecotypes and involved nsra/b mutant
861 seedlings in response to NPA/NAA treatment GSE65717 and GSE116923 (Tran
862 Vdu et al., 2016; Bazin et al., 2018), seedlings with modified expression of the
863 ASCO lincRNA GSE135376 (Rigo et al., 2020), root tip submitted to a short
864 phosphate starvation kinetic GSE128250 (Blein et al., 2020) and a lateral root
865 initiation kinetic from a binding essay of five time points without replicates (6h,
866 12h, 24h, 36h and 48h after binding, T. Blein, R. Swarup, M. Crespi and M.
867 Bennett unpublished data). All reads were quality trimmed using Trimmomatic.
868 For each library independently, cleaned reads were mapped on TAIR9 genome
869 sequence with STAR (version 2.7.2a) using Araport11 as a guided annotation
870 with the following additional parameters: --alignIntronMin 20 --alignIntronMax
871 3000. For each alignment file, StringTie (version 2.1.4) was used to predict
872 transcripts using Araport11 annotations as a guide (additional parameters: -c 2.5
873 -j 10). GFFcompare (v0.12.6) was then used to isolate the new transcripts in
874 comparison to Araport11 gene annotation (removing transcripts with class code =,
875 c, e or s). The different transcripts prediction were then combined using StringTie
876 in merge mode (additional parameters: -F 0 -T 0 -c 0 -f 0 -g 0 -i). The final set of
877 transcripts was compared against Araport11 annotation with GFFcompare
878 removing all transcripts with a class code of =, c, e, s or m. Transcripts were then
879 associated with their already annotated gene or to newly defined genes in case
880 they were predicted in unannotated portion of the genome. Coding potential was

881 then assets using COME (Hu et al., 2017), Coding Potential Calculator CPC
882 (v0.9-r2), CPC2 (Kang et al., 2017) and infernal (v1.1.2) (Nawrocki and Eddy,
883 2013) against Rfam v14.1 (Kalvari et al., 2021) with default parameters. Non-
884 coding transcripts were the ones predicted by CPC, CPC2 and COME as non-
885 coding and having no hits against tRNA, rRNA, snRNA or snoRNA genes in
886 Rfam.

887 **Integration of lincRNA gene annotations**

888 *Arabidopsis* lincRNAs were collected from public databases including Araport11
889 (Cheng et al., 2017), CANTATadb (Szczesniak et al., 2016), NONCODEv5 (Fang
890 et al., 2018), PLNlncRbase (Xuan et al., 2015), obtained from publications (Liu et
891 al., 2012; Nelson et al., 2017; Zhao et al., 2018) or shared by Andrew D. L.
892 Nelson from the Boyce Thompson Institute at Cornell using their previously
893 published method, and the in-house dataset. These collections contain lincRNA
894 annotations based on transcriptomic information coming from a wide variety of
895 organs including seedling, root, pollen, rosette leaf, endosperm, seed, siliques,
896 inflorescence, flower, floral buds, as well as abiotic stress treatments. The
897 pipeline used for the identification of putative lincRNA is described in
898 Supplemental Figure S1A. (1) LincRNA transcripts with a length of at least 200
899 bp were retained. (2) Only transcripts that were at least 500 bp away from any
900 protein-coding gene were retained and considered as intergenic (Liu et al., 2012;
901 Yamada, 2017). (3) Transcripts lacking strand information were discarded. (4)
902 The coding potential of transcripts was assessed using the Coding-Non-Coding
903 Index (CNCI; Version 2) (Sun et al., 2013), CPC2, and Pfam-scan (PFAM) (Finn
904 et al., 2016), and only those transcripts fulfilling the CPC2 (cutoff < 0), CNCI
905 (cutoff < 0) and PFAM (E-value 1e-5) criteria were retained. (6) All candidate
906 intergenic transcripts were assigned to lincRNA loci using the GFFcompare
907 program (Pertea and Pertea, 2020). The number of transcripts retained after
908 each filtering step is reported in Supplemental Figure S1B.

909 **Evolutionary conservation analysis**

910 Our set of *Arabidopsis* lincRNAs was classified into distinct evolutionary age
911 categories based on sequence similarity. The sequences of 6599 lincRNA loci
912 were extracted using BEDTools getfasta v2.30.0 (Quinlan and Hall, 2010).
913 Genome sequence data for 40 plant species were obtained from PLAZA 5.0
914 Dicots and PLAZA 5.0 Monocot (Van Bel et al., 2021), representing 26 eudicots
915 (20 rosids and 6 asterids) and 14 monocots. The 20 rosids contain 10
916 Brassicaceae species. LincRNA homologs were identified using sequence
917 similarity searches against these 40 genomes using BLASTn (Altschul et al.,
918 1990) and applying an E-value cutoff of $1e-10$ (Nelson et al., 2017).
919 Classification rules were defined to construct five evolutionary age categories. A
920 lincRNA was deemed conserved in the Angiosperm evolutionary age category
921 when at least one homolog was found in eudicots and one homolog in monocots.
922 LincRNA was defined as a eudicot/rosid-conserved lincRNA with at least one
923 homolog in rosids, one homolog in asterids, and no homolog in monocots, in
924 addition with one homolog only in rosids. The lincRNA was assigned as a
925 Brassicaceae_I_II-conserved lincRNA with at least one homolog in Brassicaceae
926 lineage I, one homolog in Brassicaceae lineage II, and no homologs outside the
927 Brassicaceae species. A lincRNA that had at least one homolog in a
928 Brassicaceae lineage I species, apart from *Arabidopsis*, was defined as
929 Brassicaceae_I-conserved lincRNA. The last category, defined as *Arabidopsis*-
930 specific lincRNA, were restricted to *Arabidopsis* lincRNAs without homologs in
931 any of the other species.

932 We downloaded the GFF annotation file from the Araport11 genome release
933 (Cheng et al., 2017) containing 27,655 protein-coding genes and 325 pri-miRNAs
934 and obtained exonic sequences. We also extracted the promoter region of 2kb
935 upstream of the transcription start site for lincRNAs, pri-miRNAs, and protein-
936 coding genes. The promoter of a gene was shortened and removed when it
937 overlapped with a nearby gene sequence.

938 The sequence conservation of exon and promoter regions per gene type was
939 evaluated using phastCons scores, which were calculated using the alignments

940 of 20 angiosperm plant genomes (Hupaló and Kern, 2013). The phastCons
941 scores were downloaded from the araTha9 genome browser available at
942 genome.genetics.rutgers.edu as a bedgraph file. The bedgraph, consisting of
943 variable width bin of equal phastCons score, was reprocessed to use a fix bin
944 width of 1nt. In case of absence of a phastCons score on a portion of the genome,
945 no bin was created which allowed making the difference between nucleotides
946 with a score (informative nucleotides) or absence of score (non-informative
947 nucleotides). The average phastCons score of exon or promoter regions was
948 computed using the BEDTools map v2.30.0 with the “-c4 -o mean” options,
949 giving the average phastCons score using only informative nucleotides. The
950 number of informative nucleotides of the genome proportion with phastCons
951 score was computed using BEDTools intersect v2.30.0. Only for loci with at least
952 50% of informative nucleotides average phastCons scores were computed and
953 reported in Figure 2 B-C (other loci were discarded).

954 **Expression analysis of lincRNAs**

955 To generate an expression atlas for lincRNAs, pri-miRNAs and protein-coding
956 genes, we used Curse (Vanechoutte and Vandepoele, 2019) to search and
957 curate relevant RNA-seq experiments. Details of the 18 RNA-seq experiments
958 across all 791 samples are shown in Supplemental Table S2 and Supplemental
959 Data Set S3. Next, we imported the expression metadata to the Prose tool
960 (Vanechoutte and Vandepoele, 2019), which downloads the raw data from SRA
961 by using the SRA toolkit, performs quality control and adapter detection,
962 trimmomatic to perform adapter clipping and quality trimming and finally Kallisto
963 (Bray et al., 2016) for quantifying transcript expression to normalized transcripts
964 per million (TPM) values. We downloaded the transcript FASTA file for 27,655
965 protein-coding genes and 325 pri-miRNAs from the Araport11 genome release
966 (Cheng et al., 2017) and retrieved transcript sequences for 6,599 lincRNAs using
967 gffread (Pertea and Pertea, 2020). Gene-level atlases were created by summing
968 the TPM values of all transcripts. We retained TPM value per gene across the

969 biological replicates and took an average of TPM values per gene from technical
970 replicates, resulting in an expression atlas covering 791 samples.

971 We used a simulation experiment (Ramskold et al., 2009; Li et al., 2016) to
972 determine thresholds for detectable expression in TPM for protein-coding genes
973 and lincRNAs. The protein-coding genes were used as true positives and the
974 lincRNAs were used as true negatives. We calculated the false-positive rate and
975 false-negative rate at different TPM thresholds for each sample. The applied
976 cutoff of 0.2 TPM to define lincRNA and Pri-miRNAs expression is a good trade-
977 off to detect lowly expressed lincRNAs and keep false-positives under control
978 (false positive rate < 0.10). A more stringent threshold TPM ≥ 2 was used to
979 define the expression of protein-coding genes.

980 To identify the clusters containing samples with related gene expression
981 information, a \log_2 -transformed gene expression matrix (TPM+1) was used for t-
982 SNE clustering with perplexity 30 and n_iter = 1000 (Pedregosa et al., 2011). 22
983 clusters were retained, the largest of which contained 113 samples and the
984 smallest contained 5 samples. Twenty-two samples were not clustered and
985 removed (Supplemental Data Set S3). Expression specificity per gene type, for
986 the gene expression matrix described above, was measured using tissue-
987 specificity index (τ), defined by the following equation (Yanai et al., 2005):

$$\tau = \frac{\sum_{i=1}^n (1 - \hat{x}_i)}{n - 1}, \hat{x}_i = \frac{x_i}{\max_{1 \leq i \leq n} (x_i)}$$

988 Where x_i was defined as the average TPM value of per cluster and n
989 corresponds to the number of clusters analyzed.

990

991 **Chromatin states analysis and TF peak annotation**

992 The enrichment of the lincRNA sets in different chromatin states (CSs) of
993 Arabidopsis from (Liu et al., 2018; Hazarika et al., 2022) was done by shuffling
994 the lincRNA genome coordinates 1000 times over the whole Arabidopsis genome.
995 Then, we compared the distribution of the number of lincRNAs expected to
996 overlap by chance with each CS with the real number of overlaps, and we used

997 these values to calculate enrichment statistics: p-value as the number of times
998 the real overlap was higher than the overlap with any of the 1000 shuffled
999 lincRNA sets, and enrichment fold as the real overlap divided by the median of
1000 overlap expected by chance (median of the 1000 shuffling events). The p-value
1001 was adjusted for multiple testing using Benjamini-Hochberg correction
1002 (significance level 0.05). For visualization purposes, the two enrichment metrics
1003 were combined into the π -value, which is the $-\log_{10}(\text{p-value}) \times \text{enrichment fold}$.
1004 TF ChIP-Seq peak coordinates were retrieved from the PlantPAN 3.0 database
1005 (Chow et al., 2019) and (Song et al., 2016). The original ChIP-Seq of KAN1,
1006 MYB44 and PIF4 were derived from these studies (Merelo et al., 2013; Pfeiffer et
1007 al., 2014; Song et al., 2016). For each peak the closest gene was identified and
1008 only peaks confirmed in two or more replicates and within a 2kb window of the
1009 gene body were retained. Starting from all TF – target gene pairs (either a
1010 protein-coding gene or a lincRNAs; Supplemental Data Set S5), enrichment
1011 analysis was performed to identify enriched TFs in different expression clusters
1012 or age categories. For all enrichment analyses the hypergeometric distribution
1013 was applied and the q-value of enrichment was determined using the Benjamini–
1014 Hochberg correction for multiple hypotheses testing. Detailed information for
1015 chromatin states, including preferential location and preferential epigenetic
1016 markers, was obtained from the Plant Chromatin State Database (PCSD) (Liu et
1017 al., 2018) and Hazarika et al., 2022.

1018 **Identification of GWAS-associated genes**

1019 GWAS data were collected from AraGWAS (Togninalli et al., 2020) and
1020 overlapped with the gene body and promoter 2kb sequences of lincRNAs to
1021 associate with the phenotype of interest. All significant associations (Permutation
1022 threshold <0.05 and FDR <0.05) were retained for screening for minor allele
1023 frequency >0.01 , resulting in 1,124 lincRNAs, covering 147 different studies.
1024 These significantly associated lincRNAs were classified into five main traits by
1025 ontological annotation, including root, seed, flower, leaf, and abiotic-related traits
1026 (see Supplemental Data Set S6).

1027 **Overlap with accessible chromatin regions**

1028 ATAC-seq data for Arabidopsis root hair, non-hair and whole roots were collected
1029 from three publications (Maher et al., 2018; Potter et al., 2018; Tannenbaum et
1030 al., 2018), and scATAC-seq data for Arabidopsis root epidermis, endodermis,
1031 stele (pericycle, xylem, phloem), and cortex cells were collected from (Dorrity et al.,
1032 2021). Cell type-specific marker peaks were identified by scATAC-seq data
1033 ($p < 0.05$ and $\text{avg_lofFC} > 0$). BEDtools intersect v2.30.0 (Quinlan and Hall, 2010)
1034 was used to detect whether TF ChIP-seq peaks associated with lincRNAs
1035 overlapped with ACR peaks. Only overlaps between ACRs and TF ChIP peaks in
1036 at least two replicates were retained, requiring that at least of 10% of the ChIP
1037 peak was covered by the ACR. BEDtools shuffle v2.30.0 (with parameter -chrom)
1038 was used to shuffle the TF ChIP-seq peaks associated with lincRNAs.

1039 **Validation of TF-lincRNA regulatory interactions using reverse**
1040 **transcription-quantitative PCR**

1041 For each replicate, total RNA from five to eight 14-day roots grown vertically on
1042 solid MS $\frac{1}{2}$ media was extracted using TRI Reagent (Sigma-Aldrich) and
1043 digested with RNase-free DNase (Fermentas) following the manufacturer's
1044 recommendations. cDNA was synthesized using Maxima Reverse Transcriptase
1045 (Thermo Scientific). Expression analysis by RT-qPCR was performed using
1046 SYBR Green master I (Roche) and the LightCycler® 96 system following a
1047 standard protocol (40 cycles, 60°C annealing). Data were analyzed using the
1048 $\Delta\Delta\text{Ct}$ method with PP2A (PROTEIN PHOSPHATASE 2A SUBUNIT A3
1049 (AT1G13320)) as reference transcript for normalization of RT-qPCR data. WT
1050 Col-0 plants grown at the same time were used as sample reference. For the
1051 dexamethasone inducible KAN1-GR expression lines analysis, after 14 days on
1052 MS media, the plants were transferred for one day either on dexamethasone
1053 containing plates (10 μ M) or only DMSO (dexamethasone solvent) as control.
1054 Primers used are listed in Supplemental Table S9. Three biological replicates
1055 were performed per condition. Statistical analyses were performed using the
1056 unpaired two-tailed Student's t-Test (GraphPad prism).

1057

1058 We use overall accuracy, precision, and recall as evaluation metrics to assess
1059 the TF-lincRNA regulatory network by RT-qPCR experiments.

$$\text{Accuracy} = \frac{TP + TN}{(TP + TN + FP + FN)}$$

$$\text{precision} = \frac{TP}{(TP + FP)}$$

$$\text{Recall} = \frac{TP}{(TP + FN)}$$

1060 Where TP is a true positive, indicating that we correctly predicted the lincRNA
1061 regulated by TF, as confirmed by significant differences in the results of RT-
1062 qPCR experiment. TN is a true negative, indicating that we predicted that a
1063 lincRNA is not regulated by TF, and the results of the RT-qPCR experiment
1064 showed no significant difference. FN is a false negative, indicating that we did not
1065 predict a lincRNA to be regulated by a specific TF, but a significant difference in
1066 the results of RT-qPCR experiments was found. Finally, a FP is a false positive,
1067 indicating we predicted that a lincRNA is regulated by TF, but there is no
1068 confirmation from the RT-qPCR experiment.

1069

1070 **Accession Numbers**

1071 Sequence data from this article can be found in the GenBank/EMBL data libraries
1072 under accession numbers_.

1073

1074 **SUPPLEMENTAL DATA**

1075 **Supplemental Figure S1.** Integration of lincRNA resources.

1076 **Supplemental Figure S2.** Definition of expressed genes and influence of
1077 sequencing depth.

1078 **Supplemental Figure S3.** Resource annotation for highly-specific expressed
1079 lincRNAs.

1080 **Supplemental Figure S4.** Regulatory properties of lincRNAs.

1081 **Supplemental Figure S5.** TF binding and expression.

1082 **Supplemental Figure S6.** RT-qPCR validation results.

1083 **Supplemental Figure S7.** TF binding for a selection of lincRNA loci confirmed by
1084 RT-qPCR experiments.

1085 **Supplemental Figure S8.** GWAS traits associated with lincRNAs.

1086 **Supplemental Table S1.** The number of homologous lincRNAs aligned to the
1087 genomes of 40 species.

1088 **Supplemental Table S2.** Summary of RNA sequencing data.

1089 **Supplemental Table S3.** List of lincRNAs with known functions.

1090 **Supplemental Table S4.** Overview of TFs ChIP-Seq data used in this study.

1091 **Supplemental Table S5.** TF and the absolute number of target genes per gene
1092 type.

1093 **Supplemental Table S6.** Overview of TFs showing significant enrichment for
1094 binding to highly-specific expressed lincRNAs for the different expression clusters
1095 (corresponds to Figure 5B).

1096 **Supplemental Table S7.** Overview of lincRNA peak annotation and RT-qPCR
1097 validation.

1098 **Supplemental Table S8.** RT-qPCR confirmed/associated with GWAS root traits
1099 lincRNAs expression information for the different samples of expression cluster

1100 19 (root) and regulatory interactions detected in ACRs (corresponds to Figure
1101 6B).

1102 **Supplemental Table S9.** RT-qPCR primer sequences used in this study.

1103 **Supplemental Data Set S1.** Full list of putative lincRNAs integrated from
1104 different resources.

1105 **Supplemental Data Set S2.** PhastCons Scores for lincRNA loci and evolutionary
1106 age category.

1107 **Supplemental Data Set S3.** Metadata for 791 samples of RNA-seq and cluster
1108 information.

1109 **Supplemental Data Set S4.** The list of the highly-specific lincRNAs.

1110 **Supplemental Data Set S5.** The table of TF-lincRNA interactions.

1111 **Supplemental Data Set S6.** The lincRNA loci containing significant GWAS hits.

1112

1113

1114 **FUNDING**

1115 The IPS2 is benefited from the support of Saclay Plant Sciences-SPS (ANR-17-
1116 EUR-0007). L.L. is supported by the China Scholarship Council for a PhD
1117 fellowship (201808530499). M.H. is supported by a PhD fellowship from the
1118 Fondation pour la Recherche Médicale (ECO202106013730).

1119

1120

1121 **ACKNOWLEDGEMENTS**

1122 We thank Mariek Dubois, Andrés Ritter, Tereza Vavrdova, Freya De Winter,
1123 Rebecca De Clercq for kindly providing the Arabidopsis lines. We thank Olivier
1124 Martin and Federico Ariel for providing comments on the manuscript.

1125

1126 **AUTHOR CONTRIBUTIONS**

1127 L.L., M.C., T.B., and K.V., designed the research. L.L., T.D., and T.B., performed
1128 data analysis. N.M.P. performed chromatin state analysis. M.H. performed
1129 experimental work. L.L., T.B., and K.V., wrote the manuscript.

1130

1131 **TABLES**

1132

1133 **FIGURE LEGENDS**

1134 **Figure 1. Overlap and gene features of Arabidopsis lincRNA annotations. (A)**

1135 Upset plot showing the intersection of lincRNA annotation in the eight resources.
1136 Each row represents a resource, reporting in parenthesis its total number of
1137 lincRNA transcripts before merging. LincRNA annotations unique to a single
1138 resource are represented as a single circle while circles connected by lines
1139 represent the intersection of lincRNA loci shared between various resources. The
1140 bar chart indicates the number of unique lincRNA loci and intersectional lincRNA
1141 loci, displaying only intersections that contain at least ten lincRNA loci. More
1142 complex overlapping patterns are not shown. **(B)** The pie chart shows the
1143 proportion of lincRNA loci supported by one or more resources. **(C)** The
1144 distribution of exon number for all lincRNA transcripts (purple), protein-coding
1145 transcripts (green), transcripts of lincRNAs supported by single resource (purple)
1146 and multiple resources (purple). Single exon and multiple exons are shown in
1147 dark and light colors, respectively. **(D)** The distribution of transcript length for
1148 lincRNAs (purple) and protein-coding genes (green).

1149

1150 **Figure 2. Sequence conservation analysis for lincRNAs of different**

1151 **evolutionary age categories. (A)** Simplified species tree reporting the number
1152 of lincRNAs found for the different age categories, which are also indicated by
1153 the grey circle sizes. Numbers in parenthesis report the number of genomes

1154 included per clade to assess sequence similarity and define a lincRNA's age
1155 category. Boxplot showing the average phastCons score for **(B)** exons and **(C)**
1156 promoter regions (2kb upstream of transcription start site) of different lincRNA
1157 age categories and gene types (lincRNAs, pri-miRNAs, protein-coding genes).
1158 The numbers in parentheses report the number of exons and promoters with at
1159 least 50% of informative nucleotides over the total number of gene bodies and
1160 promoters in that category, respectively. PhastCons score ranges from 0 (not
1161 under selection) to 1 (strong negative selection). P-values for pairwise Mann-
1162 Whitney *U* test are shown using the horizontal lines connecting the series and
1163 were corrected for multiple testing using the Benjamini-Hochberg procedure.

1164

1165 **Figure 3. Expression analysis of lincRNAs.** **(A)** Line chart showing the
1166 distribution of expression breadth for lincRNAs, pri-miRNAs and protein-coding
1167 genes across all samples. **(B)** Distribution of the maximum TPM expression
1168 levels for different lincRNA age categories and gene types (lincRNAs, pri-
1169 miRNAs and protein-coding genes). **(C)** Distribution of tissue specificity tau
1170 scores for different lincRNA age categories and gene types. Tau scores range
1171 from zero to one, where zero means widely expressed, and one means very
1172 specifically expressed (detectable in only one cluster). The black dotted line
1173 represents a tau score of 0.97. **(D)** Bar chart reporting the number of expressed
1174 and highly-specific expressed lincRNAs per expression cluster and the number of
1175 lincRNAs that are only expressed in one cluster. Cluster numbers and
1176 descriptions are shown below the chart, with numbers in parenthesis indicating
1177 the number of samples present per cluster. **(E)** Heatmap showing the proportion
1178 of highly-specific expressed lincRNAs in each cluster for each lincRNA age
1179 category. Cluster descriptions are the same as in panel D. Numbers in
1180 parenthesis report the number of genes per age category together with the
1181 fraction of lincRNAs showing highly-specific expression.

1182

1183 **Figure 4. Chromatin state and TF ChIP-Seq peak annotation for different**
1184 **gene types (A)** Dendrogram showing the enrichment for different lincRNA gene

1185 sets (x-axis) towards different chromatin states (CS) (y-axis). The values report
1186 the product of $-\log_{10}(\text{q-value})$ and the enrichment fold. Only significant
1187 enrichment values are reported (q-value < 0.05). **(B)** The proportion of peak-gene
1188 pairs present in single replicate (blue) and two or more ChIP-Seq replicates
1189 (grey). **(C)** The percentage of three gene types assigned to peaks in two or more
1190 ChIP-Seq replicates. **(D)** Distributions of the number of TF binding events for
1191 lincRNA evolutionary age categories and gene types (lincRNAs, pri-miRNAs and
1192 protein-coding genes).

1193

1194 **Figure 5. Overview of TF-lincRNA regulatory interactions in different**
1195 **expression clusters and age categories.** The dot sizes represent the number
1196 of the lincRNAs while the color represents the statistical significance. Cluster
1197 descriptions are the same as in Figure 3D. **(A)** Bubble chart showing the
1198 enrichment of TF binding for expressed lincRNAs in different expression clusters.
1199 TFs lacking significant enrichment in any of the 22 clusters are not shown. **(B)**
1200 Bubble chart showing the enrichment of TF binding for highly-specific lincRNAs in
1201 different expression clusters. TFs lacking significant enrichment in any of the 22
1202 clusters are not shown. **(C)** Bubble chart showing the enrichment of TF binding
1203 for expressed lincRNAs in different age categories.

1204

1205 **Figure 6. Experimental validation and characterization of TF-lincRNA**
1206 **regulatory interactions.** **(A)** Heatmap showing log₂ fold change (FC) of lincRNA
1207 relative expression levels in transcription factor (TF) overexpressing lines
1208 (*35S::MYB44*, *35S::PIF4* and *KAN1-GR*) or TF knockout lines (*rga28* and *pifq*
1209 (*pif1/pif2/pif3/pif4* quadruple mutants)) vs. control wild-type lines in 14-day old
1210 roots. Expression values were determined by RT-qPCR. Asterisks indicate
1211 statistically significant differences (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$) in an
1212 unpaired two-tailed Student's t-Test ($n = 3$). Solid boxes indicate TF ChIP peaks
1213 <2kb from the lincRNA gene. Dashed boxes indicate TF ChIP peaks were
1214 identified only in one ChIP-Seq replicate or the lincRNA is not the closest gene to
1215 the TF peak. **(B)** Heatmap showing cell type-specific expression of lincRNAs

1216 consistent with the expression of the regulatory TF, together with root ACR
1217 information. Yellow and green report the GWAS root-related and RT-qPCR
1218 experimentally validated lincRNA genes, respectively. The color scale represents
1219 the expression levels of lincRNAs in root cells. Triangles indicate cases where
1220 lincRNA expression and ACRs confirm regulatory interaction in the same cell
1221 type.
1222

1223 **TABLES**

1224 **Table 1. Summary of qPCR validation for TF-lincRNA gene pairs.**

1225

Line	Total number of validated lincRNAs (1)	Has a peak and is DE (TP) (2)	No peak and not DE (TN) (3)	Has a peak and is not DE (FP) (4)	No peak and is DE (FN) (5)	Has a putative peak and is DE (6)	Accuracy (7)	Precision (8)	Recall (9)
35S:MYB44	24	11	5	6	0	2	0.73	0.65	1.00
35S:PIF4	26	8	4	9	4	1	0.48	0.47	0.67
KAN1-GR	24	6	6	7	3	2	0.55	0.46	0.67
Pifq	25	8	6	8	1	2	0.61	0.50	0.89
rga-28	25	3	11	8	2	1	0.58	0.27	0.60

1226 TP = true positive, TN = true negative, FP = false positive and FN = false
 1227 negative (see Material and Methods).

1228 (1) In total 27 lincRNAs were tested in one or more lines.

1229 (2) DE refers to differential expression, indicating deregulation in a TF
 1230 perturbation line.

1231 (3) These peak-lincRNA pairs do not satisfy our stringent peak definition (present
 1232 in two or more replicates and present within 2kb of the gene body). Here a peak
 1233 is only identified in one ChIP-Seq replicate or the lincRNA is not the closest
 1234 target gene to this peak.

1235 (4) The genes present in the category “Has a putative peak and is DE” were
 1236 excluded to compute Accuracy, Precision and Recall.

1237

1238 REFERENCES

- 1239 Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local
1240 alignment search tool. *J Mol Biol* **215**, 403-410.
- 1241 Ariel, F., Lucero, L., Christ, A., Mammarella, M.F., Jegu, T., Veluchamy, A.,
1242 Mariappan, K., Latrasse, D., Blein, T., Liu, C., Benhamed, M., and Crespi, M.
1243 (2020). R-Loop Mediated trans Action of the APOLO Long Noncoding RNA. *Mol*
1244 *Cell* **77**, 1055-1065 e1054.
- 1245 Bardou, F., Ariel, F., Simpson, C.G., Romero-Barrios, N., Laporte, P., Balzergue, S.,
1246 Brown, J.W., and Crespi, M. (2014). Long noncoding RNA modulates
1247 alternative splicing regulators in Arabidopsis. *Dev Cell* **30**, 166-176.
- 1248 Bazin, J., Romero, N., Rigo, R., Charon, C., Blein, T., Ariel, F., and Crespi, M. (2018).
1249 Nuclear Speckle RNA Binding Proteins Remodel Alternative Splicing and the
1250 Non-coding Arabidopsis Transcriptome to Regulate a Cross-Talk Between Auxin
1251 and Immune Responses. *Frontiers in Plant Science* **9**.
- 1252 Beilstein, M.A., Nagalingum, N.S., Clements, M.D., Manchester, S.R., and Mathews,
1253 S. (2010). Dated molecular phylogenies indicate a Miocene origin for Arabidopsis
1254 thaliana. *Proc Natl Acad Sci U S A* **107**, 18724-18728.
- 1255 Bhogireddy, S., Mangrauthia, S.K., Kumar, R., Pandey, A.K., Singh, S., Jain, A.,
1256 Budak, H., Varshney, R.K., and Kudapa, H. (2021). Regulatory non-coding
1257 RNAs: a new frontier in regulation of plant biology. *Functional & Integrative*
1258 *Genomics* **21**, 313-330.
- 1259 Blein, T., Balzergue, C., Roule, T., Gabriel, M., Scalisi, L., Francois, T., Sorin, C.,
1260 Christ, A., Godon, C., Delannoy, E., Martin-Magniette, M.L., Nussaume, L.,
1261 Hartmann, C., Gautheret, D., Desnos, T., and Crespi, M. (2020). Landscape of
1262 the Noncoding Transcriptome Response of Two Arabidopsis Ecotypes to
1263 Phosphate Starvation. *Plant Physiol* **183**, 1058-1072.
- 1264 Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal
1265 probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-527.
- 1266 Bu, D.C., Luo, H.T., Jiao, F., Fang, S.S., Tan, C.F., Liu, Z.Y., and Zhao, Y. (2015).
1267 Evolutionary annotation of conserved long non-coding RNAs in major mammalian
1268 species. *Science China-Life Sciences* **58**, 787-798.
- 1269 Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn,
1270 J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs
1271 reveals global properties and specific subclasses. *Genes Dev* **25**, 1915-1927.
- 1272 Chen, L., Zhu, Q.H., and Kaufmann, K. (2020). Long non-coding RNAs in plants:
1273 emerging modulators of gene activity in development and stress responses.
1274 *Planta* **252**, 92.
- 1275 Chen, Q.S., Liu, K., Yu, R., Zhou, B.L., Huang, P.P., Cao, Z.X., Zhou, Y.Q., and
1276 Wang, J.H. (2021). From "Dark Matter" to "Star": Insight Into the Regulation
1277 Mechanisms of Plant Functional Long Non-Coding RNAs. *Frontiers in Plant*
1278 *Science* **12**.
- 1279 Cheng, C.Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S., and
1280 Town, C.D. (2017). Araport11: a complete reannotation of the Arabidopsis
1281 thaliana reference genome. *Plant J* **89**, 789-804.
- 1282 Chow, C.N., Lee, T.Y., Hung, Y.C., Li, G.Z., Tseng, K.C., Liu, Y.H., Kuo, P.L., Zheng,
1283 H.Q., and Chang, W.C. (2019). PlantPAN3.0: a new and updated resource for
1284 reconstructing transcriptional regulatory networks from ChIP-seq experiments in
1285 plants. *Nucleic Acids Research* **47**, D1155-D1163.

1286 **Corona-Gomez, J.A., Coss-Navarrete, E.L., Garcia-Lopez, I.J., Klapproth, C., Perez-**
1287 **Patino, J.A., and Fernandez-Valverde, S.L.** (2022). Transcriptome-guided
1288 annotation and functional classification of long non-coding RNAs in *Arabidopsis*
1289 *thaliana*. *Sci Rep* **12**, 14063.

1290 **Deng, F., Zhang, X., Wang, W., Yuan, R., and Shen, F.** (2018). Identification of
1291 *Gossypium hirsutum* long non-coding RNAs (lncRNAs) under salt stress. *BMC*
1292 *Plant Biol* **18**, 23.

1293 **Devaiah, B.N., Nagarajan, V.K., and Raghothama, K.G.** (2007). Phosphate
1294 homeostasis and root development in *Arabidopsis* are synchronized by the zinc
1295 finger transcription factor ZAT6. *Plant Physiol* **145**, 147-159.

1296 **Dorrity, M.W., Alexandre, C.M., Hamm, M.O., Vigil, A.L., Fields, S., Queitsch, C.,**
1297 **and Cuperus, J.T.** (2021). The regulatory landscape of *Arabidopsis thaliana*
1298 roots at single-cell resolution. *Nat Commun* **12**, 3334.

1299 **Fang, S.S., Zhang, L.L., Guo, J.C., Niu, Y.W., Wu, Y., Li, H., Zhao, L.H., Li, X.Y.,**
1300 **Teng, X.Y., Sun, X.H., Sun, L., Zhang, M.Q., Chen, R.S., and Zhao, Y.** (2018).
1301 NONCODEV5: a comprehensive annotation database for long non-coding RNAs.
1302 *Nucleic Acids Research* **46**, D308-D314.

1303 **Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter,**
1304 **S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J.,**
1305 **and Bateman, A.** (2016). The Pfam protein families database: towards a more
1306 sustainable future. *Nucleic Acids Res* **44**, D279-285.

1307 **Fukuda, M., Nishida, S., Kakei, Y., Shimada, Y., and Fujiwara, T.** (2019). Genome-
1308 Wide Analysis of Long Intergenic Noncoding RNAs Responding to Low-Nutrient
1309 Conditions in *Arabidopsis thaliana*: Possible Involvement of Trans-Acting siRNA3
1310 in Response to Low Nitrogen. *Plant Cell Physiol* **60**, 1961-1973.

1311 **Haudry, A., Platts, A.E., Vello, E., Hoen, D.R., Leclercq, M., Williamson, R.J.,**
1312 **Forczek, E., Joly-Lopez, Z., Steffen, J.G., Hazzouri, K.M., Dewar, K.,**
1313 **Stinchcombe, J.R., Schoen, D.J., Wang, X., Schmutz, J., Town, C.D., Edger,**
1314 **P.P., Pires, J.C., Schumaker, K.S., Jarvis, D.E., Mandakova, T., Lysak, M.A.,**
1315 **van den Bergh, E., Schranz, M.E., Harrison, P.M., Moses, A.M., Bureau, T.E.,**
1316 **Wright, S.I., and Blanchette, M.** (2013). An atlas of over 90,000 conserved
1317 noncoding sequences provides insight into crucifer regulatory regions. *Nat Genet*
1318 **45**, 891-898.

1319 **Hawker, N.P., and Bowman, J.L.** (2004). Roles for Class III HD-Zip and KANADI genes
1320 in *Arabidopsis* root development. *Plant Physiol* **135**, 2261-2270.

1321 **Hazarika, R.R., Serra, M., Zhang, Z., Zhang, Y., Schmitz, R.J., and Johannes, F.**
1322 (2022). Molecular properties of epimutation hotspots. *Nat Plants* **8**, 146-156.

1323 **He, H., Zhou, Y.F., Yang, Y.W., Zhang, Z., Lei, M.Q., Feng, Y.Z., Zhang, Y.C., Chen,**
1324 **Y.Q., Lian, J.P., and Yu, Y.** (2021). Genome-Wide Analysis Identified a Set of
1325 Conserved lncRNAs Associated with Domestication-Related Traits in Rice.
1326 *International Journal of Molecular Sciences* **22**.

1327 **Heo, J.B., and Sung, S.** (2011). Vernalization-mediated epigenetic silencing by a long
1328 intronic noncoding RNA. *Science* **331**, 76-79.

1329 **Heyndrickx, K.S., Van de Velde, J., Wang, C., Weigel, D., and Vandepoele, K.** (2014).
1330 A functional and evolutionary perspective on transcription factor binding in
1331 *Arabidopsis thaliana*. *Plant Cell* **26**, 3894-3910.

1332 **Hu, L., Xu, Z., Hu, B., and Lu, Z.J.** (2017). COME: a robust coding potential calculation
1333 tool for lncRNA identification and characterization based on multiple features.
1334 *Nucleic Acids Res* **45**, e2.

1335 Huang, Y., Sicar, S., Ramirez-Prado, J.S., Manza-Mianza, D., Antunez-Sanchez, J.,
1336 Brik-Chaouche, R., Rodriguez-Granados, N.Y., An, J., Bergounioux, C.,
1337 Mahfouz, M.M., Hirt, H., Crespi, M., Concia, L., Barneche, F., Amiard, S.,
1338 Probst, A.V., Gutierrez-Marcos, J., Ariel, F., Raynaud, C., Latrasse, D., and
1339 Benhamed, M. (2021). Polycomb-dependent differential chromatin
1340 compartmentalization determines gene coregulation in Arabidopsis. *Genome*
1341 *Research* **31**, 1230-+.

1342 Hupaló, D., and Kern, A.D. (2013). Conservation and functional element discovery in 20
1343 angiosperm plant genomes. *Mol Biol Evol* **30**, 1729-1744.

1344 Jha, U.C., Nayyar, H., Jha, R., Khurshid, M., Zhou, M., Mantri, N., and Siddique,
1345 K.H.M. (2020). Long non-coding RNAs: emerging players regulating plant abiotic
1346 stress response and adaptation. *BMC Plant Biol* **20**, 466.

1347 Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. (2006). MicroRNAs and their
1348 regulatory roles in plants. *Annual Review of Plant Biology* **57**, 19-53.

1349 Kalvari, I., Nawrocki, E.P., Ontiveros-Palacios, N., Argasinska, J., Lamkiewicz, K.,
1350 Marz, M., Griffiths-Jones, S., Toffano-Nioche, C., Gautheret, D., Weinberg, Z.,
1351 Rivas, E., Eddy, S.R., Finn, R.D., Bateman, A., and Petrov, A.I. (2021). Rfam
1352 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic*
1353 *Acids Res* **49**, D192-D200.

1354 Kang, Y.J., Yang, D.C., Kong, L., Hou, M., Meng, Y.Q., Wei, L., and Gao, G. (2017).
1355 CPC2: a fast and accurate coding potential calculator based on sequence
1356 intrinsic features. *Nucleic Acids Res* **45**, W12-W16.

1357 Ke, L.L., Zhou, Z.W., Xu, X.W., Wang, X., Liu, Y.L., Xu, Y.T., Huang, Y., Wang, S.T.,
1358 Deng, X.X., Chen, L.L., and Xu, Q. (2019). Evolutionary dynamics of lincRNA
1359 transcription in nine citrus species. *Plant Journal* **98**, 912-927.

1360 Kim, D.H., and Sung, S. (2017). Vernalization-Triggered Intragenic Chromatin Loop
1361 Formation by Long Noncoding RNAs. *Dev Cell* **40**, 302-312 e304.

1362 Kindgren, P., Ard, R., Ivanov, M., and Marquardt, S. (2018). Transcriptional read-
1363 through of the long non-coding RNA SVALKKA governs plant cold acclimation. *Nat*
1364 *Commun* **9**, 4561.

1365 Kramer, M.C., Kim, H.J., Palos, K.R., Garcia, B.A., Lyons, E., Beilstein, M.A., Nelson,
1366 A.D.L., and Gregory, B.D. (2022). A Conserved Long Intergenic Non-coding
1367 RNA Containing snoRNA Sequences, IncCOBRA1, Affects Arabidopsis
1368 Germination and Development. *Front Plant Sci* **13**, 906603.

1369 Li, Q.Q., Zhang, Z., Zhang, C.X., Wang, Y.L., Liu, C.B., Wu, J.C., Han, M.L., Wang,
1370 Q.X., and Chao, D.Y. (2022). Phytochrome-interacting factors orchestrate
1371 hypocotyl adventitious root initiation in Arabidopsis. *Development* **149**.

1372 Li, S., Yamada, M., Han, X., Ohler, U., and Benfey, P.N. (2016). High-Resolution
1373 Expression Map of the Arabidopsis Root Reveals Alternative Splicing and
1374 lincRNA Regulation. *Dev Cell* **39**, 508-522.

1375 Li, S.X., Yu, X., Lei, N., Cheng, Z.H., Zhao, P.J., He, Y.K., Wang, W.Q., and Peng, M.
1376 (2017). Genome-wide identification and functional prediction of cold and/or
1377 drought-responsive lincRNAs in cassava. *Scientific Reports* **7**.

1378 Liu, F., Marquardt, S., Lister, C., Swiezewski, S., and Dean, C. (2010). Targeted 3'
1379 processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing.
1380 *Science* **327**, 94-97.

1381 Liu, J., Wang, H., and Chua, N.H. (2015). Long noncoding RNA transcriptome of plants.
1382 *Plant Biotechnol J* **13**, 319-328.

1383 **Liu, J., Jung, C., Xu, J., Wang, H., Deng, S.L., Bernad, L., Arenas-Huertero, C., and**
1384 **Chua, N.H.** (2012). Genome-Wide Analysis Uncovers Regulation of Long
1385 Intergenic Noncoding RNAs in Arabidopsis. *Plant Cell* **24**, 4333-4345.

1386 **Liu, Y., Tian, T., Zhang, K., You, Q., Yan, H., Zhao, N., Yi, X., Xu, W., and Su, Z.**
1387 (2018). PCSD: a plant chromatin state database. *Nucleic Acids Res* **46**, D1157-
1388 D1167.

1389 **Lucero, L., Ferrero, L., Fonouni-Farde, C., and Ariel, F.** (2021). Functional
1390 classification of plant long noncoding RNAs: a transcript is known by the
1391 company it keeps. *New Phytologist* **229**, 1251-1260.

1392 **Ma, J.C., Bai, X.T., Luo, W.C., Feng, Y.N., Shao, X.M., Bai, Q.X., Sun, S.J., Long,**
1393 **Q.M., and Wan, D.S.** (2019). Genome-Wide Identification of Long Noncoding
1394 RNAs and Their Responses to Salt Stress in Two Closely Related Poplars.
1395 *Frontiers in Genetics* **10**.

1396 **Maher, K.A., Bajic, M., Kajala, K., Reynoso, M., Pauluzzi, G., West, D.A., Zumstein,**
1397 **K., Woodhouse, M., Bubb, K., Dorrity, M.W., Queitsch, C., Bailey-Serres, J.,**
1398 **Sinha, N., Brady, S.M., and Deal, R.B.** (2018). Profiling of Accessible Chromatin
1399 Regions across Multiple Plant Species and Cell Types Reveals Common Gene
1400 Regulatory Principles and New Control Modules. *Plant Cell* **30**, 15-36.

1401 **Mattick, J.S., Amaral, P.P., Carninci, P., Carpenter, S., Chang, H.Y., Chen, L.L.,**
1402 **Chen, R., Dean, C., Dinger, M.E., Fitzgerald, K.A., Gingeras, T.R., Guttman,**
1403 **M., Hirose, T., Huarte, M., Johnson, R., Kanduri, C., Kapranov, P., Lawrence,**
1404 **J.B., Lee, J.T., Mendell, J.T., Mercer, T.R., Moore, K.J., Nakagawa, S., Rinn,**
1405 **J.L., Spector, D.L., Ulitsky, I., Wan, Y., Wilusz, J.E., and Wu, M.** (2023). Long
1406 non-coding RNAs: definitions, functions, challenges and recommendations. *Nat*
1407 *Rev Mol Cell Biol*.

1408 **Mattioli, K., Volders, P.J., Gerhardinger, C., Lee, J.C., Maass, P.G., Mele, M., and**
1409 **Rinn, J.L.** (2019). High-throughput functional analysis of lncRNA core promoters
1410 elucidates rules governing tissue specificity. *Genome Research* **29**, 344-355.

1411 **Merelo, P., Xie, Y.K., Brand, L., Ott, F., Weigel, D., Bowman, J.L., Heisler, M.G., and**
1412 **Wenkel, S.** (2013). Genome-Wide Identification of KANADI1 Target Genes. *Plos*
1413 *One* **8**.

1414 **Mohammadin, S., Edger, P.P., Pires, J.C., and Schranz, M.E.** (2015). Positionally-
1415 conserved but sequence-diverged: identification of long non-coding RNAs in the
1416 Brassicaceae and Cleomaceae. *BMC Plant Biol* **15**, 217.

1417 **Moison, M., Pacheco, J.M., Lucero, L., Fonouni-Farde, C., Rodriguez-Melo, J.,**
1418 **Mansilla, N., Christ, A., Bazin, J., Benhamed, M., Ibanez, F., Crespi, M.,**
1419 **Estevez, J.M., and Ariel, F.** (2021). The lncRNA APOLO interacts with the
1420 transcription factor WRKY42 to trigger root hair cell expansion in response to
1421 cold. *Mol Plant* **14**, 937-948.

1422 **Moubayidin, L., Salvi, E., Giustini, L., Terpstra, I., Heidstra, R., Costantino, P., and**
1423 **Sabatini, S.** (2016). A SCARECROW-based regulatory circuit controls
1424 Arabidopsis thaliana meristem size from the root endodermis. *Planta* **243**, 1159-
1425 1168.

1426 **Nawrocki, E.P., and Eddy, S.R.** (2013). Infernal 1.1: 100-fold faster RNA homology
1427 searches. *Bioinformatics* **29**, 2933-2935.

1428 **Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., Baker,**
1429 **J.C., Grutzner, F., and Kaessmann, H.** (2014). The evolution of lncRNA
1430 repertoires and expression patterns in tetrapods. *Nature* **505**, 635-+.

1431 **Nelson, A.D.L., Devisetty, U.K., Palos, K., Haug-Baltzell, A.K., Lyons, E., and**
1432 **Beilstein, M.A.** (2017). Evolinc:A Tool for the Identification and Evolutionary
1433 Comparison of Long Intergenic Non-coding RNAs. *Frontiers in Genetics* **8**.
1434 **Nelson, A.D.L., Forsythe, E.S., Devisetty, U.K., Clausen, D.S., Haug-Batzell, A.K.,**
1435 **Meldrum, A.M.R., Frank, M.R., Lyons, E., and Beilstein, M.A.** (2016). A
1436 Genomic Analysis of Factors Driving lincRNA Diversification: Lessons from
1437 Plants. *G3-Genes Genomes Genetics* **6**, 2881-2891.
1438 **O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli,**
1439 **M., Gallavotti, A., and Ecker, J.R.** (2016). Cistrome and Epicistrome Features
1440 Shape the Regulatory DNA Landscape. *Cell* **166**, 1598.
1441 **O'Maoileidigh, D.S., Graciet, E., and Wellmer, F.** (2014). Gene networks controlling
1442 *Arabidopsis thaliana* flower development. *New Phytol* **201**, 16-30.
1443 **Pacheco, J.M., Mansilla, N., Moison, M., Lucero, L., Gabarain, V.B., Ariel, F., and**
1444 **Estevez, J.M.** (2021). The lincRNA APOLO and the transcription factor WRKY42
1445 target common cell wall EXTENSIN encoding genes to trigger root hair cell
1446 elongation. *Plant Signal Behav* **16**, 1920191.
1447 **Palos, K., Dittrich, A.C.N., Yu, L.A., Brock, J.R., Railey, C.E., Wu, H.Y.L.,**
1448 **Sokolowska, E., Skirycz, A., Hsu, P.Y., Gregory, B.D., Lyons, E., Beilstein,**
1449 **M.A., and Nelson, A.D.L.** (2022). Identification and functional annotation of long
1450 intergenic non-coding RNAs in Brassicaceae. *Plant Cell*.
1451 **Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,**
1452 **Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos,**
1453 **A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.** (2011).
1454 Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*
1455 **12**, 2825-2830.
1456 **Perte, G., and Perte, M.** (2020). GFF Utilities: GffRead and GffCompare. *F1000Res* **9**.
1457 **Pfeiffer, A., Shi, H., Tepperman, J.M., Zhang, Y., and Quail, P.H.** (2014).
1458 Combinatorial Complexity in a Transcriptionally Centered Signaling Hub in
1459 *Arabidopsis*. *Molecular Plant* **7**, 1598-1618.
1460 **Ponting, C.P., and Haerty, W.** (2022). Genome-Wide Analysis of Human Long
1461 Noncoding RNAs: A Provocative Review. *Annu Rev Genomics Hum Genet*.
1462 **Potter, K.C., Wang, J., Schaller, G.E., and Kieber, J.J.** (2018). Cytokinin modulates
1463 context-dependent chromatin accessibility through the type-B response
1464 regulators. *Nat Plants* **4**, 1102-1111.
1465 **Qi, X., Xie, S., Liu, Y., Yi, F., and Yu, J.** (2013). Genome-wide annotation of genes and
1466 noncoding RNAs of foxtail millet in response to simulated drought stress by deep
1467 sequencing. *Plant Mol Biol* **83**, 459-473.
1468 **Qin, T., Zhao, H., Cui, P., Albeshar, N., and Xiong, L.** (2017). A Nucleus-Localized
1469 Long Non-Coding RNA Enhances Drought and Salt Stress Tolerance. *Plant*
1470 *Physiol* **175**, 1321-1336.
1471 **Quinlan, A.R., and Hall, I.M.** (2010). BEDTools: a flexible suite of utilities for comparing
1472 genomic features. *Bioinformatics* **26**, 841-842.
1473 **Rai, M.I., Alam, M., Lightfoot, D.A., Gurha, P., and Afzal, A.J.** (2019). Classification
1474 and experimental identification of plant long non-coding RNAs. *Genomics* **111**,
1475 997-1005.
1476 **Ramskold, D., Wang, E.T., Burge, C.B., and Sandberg, R.** (2009). An abundance of
1477 ubiquitously expressed genes revealed by tissue transcriptome sequence data.
1478 *PLoS Comput Biol* **5**, e1000598.
1479 **Ransbotyn, V., Yeger-Lotem, E., Basha, O., Acuna, T., Verduyn, C., Gordon, M.,**
1480 **Chalifa-Caspi, V., Hannah, M.A., and Barak, S.** (2015). A combination of gene

1481 expression ranking and co-expression network analysis increases discovery rate
1482 in large-scale mutant screens for novel *Arabidopsis thaliana* abiotic stress genes.
1483 *Plant Biotechnol J* **13**, 501-513.

1484 **Ransohoff, J.D., Wei, Y., and Khavari, P.A.** (2018). The functions and unique features
1485 of long intergenic non-coding RNA. *Nat Rev Mol Cell Biol* **19**, 143-157.

1486 **Rigo, R., Bazin, J., Romero-Barrios, N., Moison, M., Lucero, L., Christ, A.,**
1487 **Benhamed, M., Blein, T., Huguet, S., Charon, C., Crespi, M., and Ariel, F.**
1488 (2020). The *Arabidopsis* lncRNA ASCO modulates the transcriptome through
1489 interaction with splicing factors. *EMBO Rep* **21**, e48977.

1490 **Roule, T., Crespi, M., and Blein, T.** (2022a). Regulatory long non-coding RNAs in root
1491 growth and development. *Biochem Soc Trans* **50**, 403-412.

1492 **Roule, T., Christ, A., Hussain, N., Huang, Y., Hartmann, C., Benhamed, M.,**
1493 **Gutierrez-Marcos, J., Ariel, F., Crespi, M., and Blein, T.** (2022b). The lncRNA
1494 MARS modulates the epigenetic reprogramming of the maternal cluster in
1495 response to ABA. *Mol Plant* **15**, 840-856.

1496 **Sanchita, Trivedi, P.K., and Asif, M.H.** (2020). Updates on plant long non-coding RNAs
1497 (lncRNAs): the regulatory components. *Plant Cell Tissue and Organ Culture* **140**,
1498 259-269.

1499 **Sarropoulos, I., Marin, R., Cardoso-Moreira, M., and Kaessmann, H.** (2019).
1500 Developmental dynamics of lncRNAs across mammalian organs and species.
1501 *Nature* **571**, 510-+.

1502 **Schmitz, R.J., Grotewold, E., and Stam, M.** (2022). Cis-regulatory sequences in plants:
1503 Their importance, discovery, and future challenges. *Plant Cell* **34**, 718-741.

1504 **Severing, E., Faino, L., Jamge, S., Busscher, M., Kuijer-Zhang, Y., Bellinazzo, F.,**
1505 **Busscher-Lange, J., Fernandez, V., Angenent, G.C., Immink, R.G.H., and**
1506 **Pajoro, A.** (2018). *Arabidopsis thaliana* ambient temperature responsive
1507 lncRNAs. *BMC Plant Biol* **18**, 145.

1508 **Shafiq, S., Li, J.R., and Sun, Q.W.** (2016). Functions of plants long non-coding RNAs.
1509 *Biochimica Et Biophysica Acta- Gene Regulatory Mechanisms* **1859**, 155-162.

1510 **Shea, D.J., Nishida, N., Takada, S., Itabashi, E., Takahashi, S., Akter, A., Miyaji, N.,**
1511 **Osabe, K., Mehraj, H., Shimizu, M., Seki, M., Kakizaki, T., Okazaki, K.,**
1512 **Dennis, E.S., and Fujimoto, R.** (2019). Long noncoding RNAs in *Brassica rapa*
1513 *L.* following vernalization. *Sci Rep* **9**, 9302.

1514 **Shuai, P., Liang, D., Tang, S., Zhang, Z.J., Ye, C.Y., Su, Y.Y., Xia, X.L., and Yin, W.L.**
1515 (2014). Genome-wide identification and functional prediction of novel and
1516 drought-responsive lincRNAs in *Populus trichocarpa*. *Journal of Experimental*
1517 *Botany* **65**, 4975-4983.

1518 **Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K.,**
1519 **Clawson, H., Spieth, J., Hillier, L.W., Richards, S., Weinstock, G.M., Wilson,**
1520 **R.K., Gibbs, R.A., Kent, W.J., Miller, W., and Haussler, D.** (2005).
1521 Evolutionarily conserved elements in vertebrate, insect, worm, and yeast
1522 genomes. *Genome Res* **15**, 1034-1050.

1523 **Song, L., Huang, S.S.C., Wise, A., Castanon, R., Nery, J.R., Chen, H.M., Watanabe,**
1524 **M., Thomas, J., Bar-Joseph, Z., and Ecker, J.R.** (2016). A transcription factor
1525 hierarchy defines an environmental stress response network. *Science* **354**.

1526 **Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., Liu, Y., Chen, R., and Zhao, Y.**
1527 (2013). Utilizing sequence intrinsic composition to classify protein-coding and
1528 long non-coding transcripts. *Nucleic Acids Research* **41**.

- 1529 Sun, Z., Nair, A., Chen, X., Prodduturi, N., Wang, J., and Kocher, J.P. (2017).
 1530 UCInCR: Ultrafast and comprehensive long non-coding RNA detection from RNA-
 1531 seq. *Sci Rep* **7**, 14196.
- 1532 Szczesniak, M.W., Rosikiewicz, W., and Makalowska, I. (2016). CANTATAdb: A
 1533 Collection of Plant Long Non-Coding RNAs. *Plant Cell Physiol* **57**, e8.
- 1534 Szczesniak, M.W., Kubiak, M.R., Wanowska, E., and Makalowska, I. (2021).
 1535 Comparative genomics in the search for conserved long noncoding RNAs.
 1536 *Essays Biochem* **65**, 741-749.
- 1537 Tannenbaum, M., Sarusi-Portuguez, A., Krispil, R., Schwartz, M., Loza, O.,
 1538 Benichou, J.I.C., Mosquna, A., and Hakim, O. (2018). Regulatory chromatin
 1539 landscape in *Arabidopsis thaliana* roots uncovered by coupling INTACT and
 1540 ATAC-seq. *Plant Methods* **14**, 113.
- 1541 Togninalli, M., Seren, U., Freudenthal, J.A., Monroe, J.G., Meng, D., Nordborg, M.,
 1542 Weigel, D., Borgwardt, K., Korte, A., and Grimm, D.G. (2020). AraPheno and
 1543 the AraGWAS Catalog 2020: a major database update including RNA-Seq and
 1544 knockout mutation data for *Arabidopsis thaliana*. *Nucleic Acids Res* **48**, D1063-
 1545 D1068.
- 1546 Tominaga-Wada, R., and Wada, T. (2016). The ectopic localization of CAPRICE LIKE
 1547 MYB3 protein in *Arabidopsis* root epidermis. *J Plant Physiol* **199**, 111-115.
- 1548 Tran Vdu, T., Souiai, O., Romero-Barrios, N., Crespi, M., and Gautheret, D. (2016).
 1549 Detection of generic differential RNA processing events from RNA-seq data. *RNA*
 1550 *Biol* **13**, 59-67.
- 1551 Ulitsky, I. (2016). Evolution to the rescue: using comparative genomics to understand
 1552 long non-coding RNAs. *Nature Reviews Genetics* **17**, 601-614.
- 1553 Van Bel, M., Silvestri, F., Weitz, E.M., Kreft, L., Botzki, A., Coppens, F., and
 1554 Vandepoele, K. (2021). PLAZA 5.0: extending the scope and power of
 1555 comparative and functional genomics in plants. *Nucleic Acids Res*.
- 1556 Van de Velde, J., Van Bel, M., Vanechoutte, D., and Vandepoele, K. (2016). A
 1557 Collection of Conserved Noncoding Sequences to Study Gene Regulation in
 1558 Flowering Plants. *Plant Physiol* **171**, 2586-2598.
- 1559 Vanechoutte, D., and Vandepoele, K. (2019). Curse: building expression atlases and
 1560 co-expression networks from public RNA-Seq data. *Bioinformatics* **35**, 2880-2881.
- 1561 Wang, H., Niu, Q.W., Wu, H.W., Liu, J., Ye, J., Yu, N., and Chua, N.H. (2015). Analysis
 1562 of non-coding transcriptome in rice and maize uncovers roles of conserved
 1563 lncRNAs associated with agriculture traits. *Plant Journal* **84**, 404-416.
- 1564 Wu, H., Yang, L., and Chen, L.L. (2017). The Diversity of Long Noncoding RNAs and
 1565 Their Generation. *Trends in Genetics* **33**, 540-552.
- 1566 Xuan, H.D., Zhang, L.Z., Liu, X.S., Han, G.M., Li, J., Li, X., Liu, A.G., Liao, M.Z., and
 1567 Zhang, S.H. (2015). PLNlncRbase: A resource for experimentally identified
 1568 lncRNAs in plants. *Gene* **573**, 328-332.
- 1569 Yamada, M. (2017). Functions of long intergenic non-coding (linc) RNAs in plants. *J*
 1570 *Plant Res* **130**, 67-73.
- 1571 Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-
 1572 Even, A., Horn-Saban, S., Safran, M., Domany, E., Lancet, D., and Shmueli,
 1573 O. (2005). Genome-wide midrange transcription profiles reveal expression level
 1574 relationships in human tissue specification. *Bioinformatics* **21**, 650-659.
- 1575 Yu, C.P., Lin, J.J., and Li, W.H. (2016). Positional distribution of transcription factor
 1576 binding sites in *Arabidopsis thaliana*. *Sci Rep* **6**, 25164.
- 1577 Zhang, L., Chen, F., Zhang, X., Li, Z., Zhao, Y., Lohaus, R., Chang, X., Dong, W., Ho,
 1578 S.Y.W., Liu, X., Song, A., Chen, J., Guo, W., Wang, Z., Zhuang, Y., Wang, H.,

1579 Chen, X., Hu, J., Liu, Y., Qin, Y., Wang, K., Dong, S., Liu, Y., Zhang, S., Yu, X.,
1580 Wu, Q., Wang, L., Yan, X., Jiao, Y., Kong, H., Zhou, X., Yu, C., Chen, Y., Li, F.,
1581 Wang, J., Chen, W., Chen, X., Jia, Q., Zhang, C., Jiang, Y., Zhang, W., Liu, G.,
1582 Fu, J., Chen, F., Ma, H., Van de Peer, Y., and Tang, H. (2020). The water lily
1583 genome and the early evolution of flowering plants. *Nature* **577**, 79-84.
1584 **Zhao, Q., Li, M., Jia, Z., Liu, F., Ma, H., Huang, Y., and Song, S.** (2016). AtMYB44
1585 Positively Regulates the Enhanced Elongation of Primary Roots Induced by N-3-
1586 Oxo-Hexanoyl-Homoserine Lactone in *Arabidopsis thaliana*. *Mol Plant Microbe*
1587 Interact **29**, 774-785.
1588 **Zhao, X.Y., Li, J.R., Lian, B., Gu, H.Q., Li, Y., and Qi, Y.J.** (2018). Global identification
1589 of *Arabidopsis* lncRNAs reveals the regulation of MAF4 by a natural antisense
1590 RNA. *Nature Communications* **9**.
1591