# Regulatory and evolutionary annotation identifies KAN1, MYB44 and PIF4 as regulators of Arabidopsis lincRNAs expressed in root

Li Liu[1,2], Michel Heidecker[3,4], Thomas Depuydt[1,2], Nicolas Manosalva Perez[1,2], Martin Crespi[3,4], Thomas Blein[3,4*], Klaas Vandepoele[1,2,4*]

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

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

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

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

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

* Shared last author

Corresponding author: klaas.vandepoele@psb.vib-ugent.be

Running title: TF regulatory annotation of Arabidopsis root lincRNAs

**ABSTRACT**

Thousands of long intergenic noncoding RNAs (lincRNAs) have been identified in plant genomes. While some lincRNAs have been characterized as important regulators in different biological processes, little is known about the transcriptional regulation for most plant lincRNAs. Through the integration of eight annotation resources, we defined 6,599 high-confidence lincRNA loci in *Arabidopsis thaliana*. For lincRNAs belonging to different evolutionary age categories, we identified major differences in sequence and chromatin features, as well as in the level of conservation and selection. Spatiotemporal gene expression profiles combined with transcription factor (TF) chromatin immunoprecipitation data was used to construct a TF-lincRNA regulatory network containing 2,659 lincRNAs and 43,233 interactions. We experimentally confirmed the role of three TFs, KAN1, MYB44, and PIF4, as key regulators controlling root-specific lincRNA expression, demonstrating the predictive power of our network. Furthermore, we identified 58 lincRNAs regulated by these TFs showing strong root cell-type specific expression and that are linked with GWAS genetic associations related to root system development and growth. The multi-omics approach applied in this study sheds light on the global regulatory complexity of plant lincRNA networks and pinpoints a role of specific TFs in lincRNA regulation in roots.

## Introduction

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

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

mediated flowering time (Severing et al., 2018). Arabidopsis lateral root development is regulated by the *Alternative Splicing COmpetitor* (*ASCO*) lincRNA, which modulates AS by interacting with the multiple splicing factors (Bardou et al., 2014; Rigo et al., 2020), and the *AUXIN-REGULATED PROMOTER LOOP* (*APOLO*) lincRNA, which influences the local chromatin conformation and the activity of several Auxin-Responsive genes (Ariel et al., 2020). Plants lacking *CONSERVED INBRASSICA RAPA1* (*lncCOBRA1*) were found to show a delayed germination and were overall smaller than wild-type plants (Kramer et al., 2022). Many lincRNAs are differentially expressed in various stress responses, including drought (Qi et al., 2013; Shuai et al., 2014; Li et al., 2017; Qin et al., 2017), cold (Li et al., 2017; Zhao et al., 2018; Shea et al., 2019), salinity (Deng et al., 2018; Fukuda et al., 2019), and nutrient deficiency (Fukuda et al., 2019), implying that lincRNAs may be involved in plant stress resilience (Jha et al., 2020). Interestingly, some of the confirmed functional lincRNAs interact with transcription factors (TFs) to activate or repress the expression of associated genes, such as *APOLO* that interacts with WRKY42 to form a regulatory hub that controls the activity of *RHD6* and promotes the expansion of root hair cell at low temperatures (Moison et al., 2021; Pacheco et al., 2021).

Recently, a lot of attention has been placed on the evolutionary conservation of lincRNAs, which is generally associated with functionality (Ulitsky, 2016; Szczesniak et al., 2021). The conservation of noncoding transcripts can be considered at the level of the primary sequence, position, splice sites, RNA structure, and transcriptional level (Ulitsky, 2016; Szczesniak et al., 2021). However, most lincRNA sequences undergo rapid evolution and are poorly conserved (Ransohoff et al., 2018). In a study by Wang et al., only 5% of 2,281 rice lincRNAs had sequence similarity to maize lincRNAs. It was also found that the positional conservation of lincRNAs was much higher than their sequence conservation (Wang et al., 2015b). Nelson and co-workers reported that 22% of 1180 Arabidopsis lincRNA loci were conserved in 10 Brassicaceae genomes (Nelson et al., 2016). These conserved lincRNAs exhibited higher expression levels, stress-responsiveness and their gene body overlapped with conserved

107 noncoding sequences (CNSs), suggesting a role of their conserved sequence in
108 a genomic context (Nelson et al., 2016).
109 While different studies have reported on the identification and expression of
110 lincRNAs in plants (Wang et al., 2015a; Nelson et al., 2016; Ke et al., 2019; He et
111 al., 2021), a comprehensive overview of the different genomic features
112 contributing to the expression, regulation and evolutionary conservation of plant
113 lincRNAs is missing. How lincRNAs are embedded in transcriptional networks
114 controlling different biological processes remains largely unknown. Furthermore,
115 prioritizing lincRNAs for downstream functional analysis is not straightforward
116 without knowing the regulatory network where they are involved in. Here, we
117 integrated different Arabidopsis lincRNAs gene annotations and explored various
118 functional genomics datasets to characterize lincRNA expression in a biologically
119 relevant context. We leveraged large-scale expression datasets and protein-DNA
120 interaction data to study the molecular networks controlling lincRNA gene activity.
121 Combined with evolutionary conservation analysis, we explored the global
122 transcriptional regulatory properties of different evolutionary age categories and,
123 through regulatory network analysis, identified specific TFs controlling lincRNA
124 regulation in roots.
125
126

127 **RESULTS**

128 **Integration of lincRNA annotations in *Arabidopsis***

129 A substantial number of lncRNA transcripts in *Arabidopsis* have been identified
130 and several publicly available resources for the annotation of lncRNAs have been
131 developed (Jha et al., 2020). In contrast to antisense lncRNAs which generally
132 regulate the overlapping coding gene, much less is known about the potential
133 targets of lincRNAs, so we focus our study on these transcripts. Indeed,
134 lincRNAs transcription and evolution are independent of the surrounding genes,
135 in contrast to antisense lncRNAs that are constrained by the coding genes they
136 overlap with. To integrate and unify previously identified lincRNAs, annotations
137 were collected from different resources, including Araport11 (Cheng et al., 2017),

138 CANTATAdb (Szczesniak et al., 2016), NONCODEv5 (Fang et al., 2018),
139 PLNlncRbase (Xuan et al., 2015), key lncRNA research articles (Liu et al., 2012;
140 Nelson et al., 2017; Zhao et al., 2018), and new predictions based on root related
141 stranded RNA-seq (Material and Methods). Next, a pipeline was designed to
142 define a unified set of high-confidence lincRNAs by discarding transcripts with
143 length below 200 bp, removing transcripts that overlapped with protein-coding
144 genes (antisense lncRNAs), re-evaluating the coding potential of the transcripts,
145 and merging the remaining transcripts from various resources (see Materials and
146 Methods, Supplemental Figure S1A-B). In total, we identified 7,706 lincRNA
147 transcripts covering 6,599 high-confidence lincRNA loci (see Supplemental Data
148 Set S1). To explore the overlap of this high-confidence lincRNA gene set with the
149 individual annotations, we assessed the overlap between the different resources
150 (Figure 1A). A total of 4,955 (75.1%) lincRNA loci were supported by only one
151 resource and the remaining 1,644 (24.9%) lincRNA loci were derived from two or
152 more resources (Figure 1B). Araport11 contained the highest number of shared
153 loci and Liu et al. (2012) contained the highest number of unique loci (Chen et al.,
154 2021). Next, the genomic features of the high-confidence lincRNA transcripts
155 were compared to those of protein-coding transcripts. 6,428 (83%) lincRNA
156 transcripts and 6,250 (13%) protein-coding transcripts contained single exons,
157 while 1,278 (17%) lincRNA transcripts and 42,109 (87%) protein-coding
158 transcripts contained multiple exons. Furthermore, a higher frequency of multi-
159 exon transcripts was found in lincRNAs supported by two or more resources (974,
160 36%) than in those supported by a unique resource (304, 6%) (Figure 1C). The
161 transcript length distribution for lincRNAs showed a U-shape curve with the
162 majority of transcripts being 200-300bp long (Figure 1D).

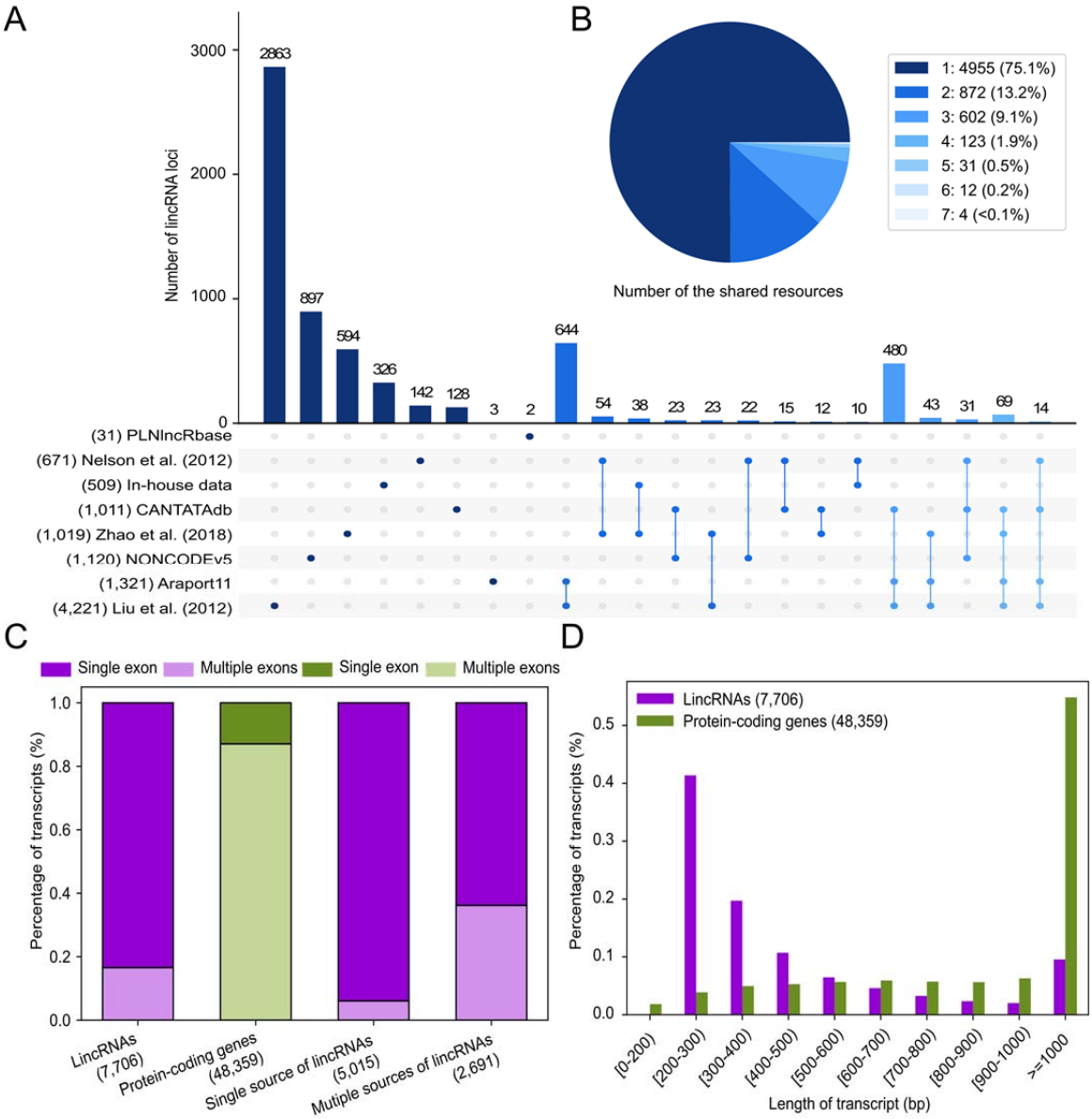


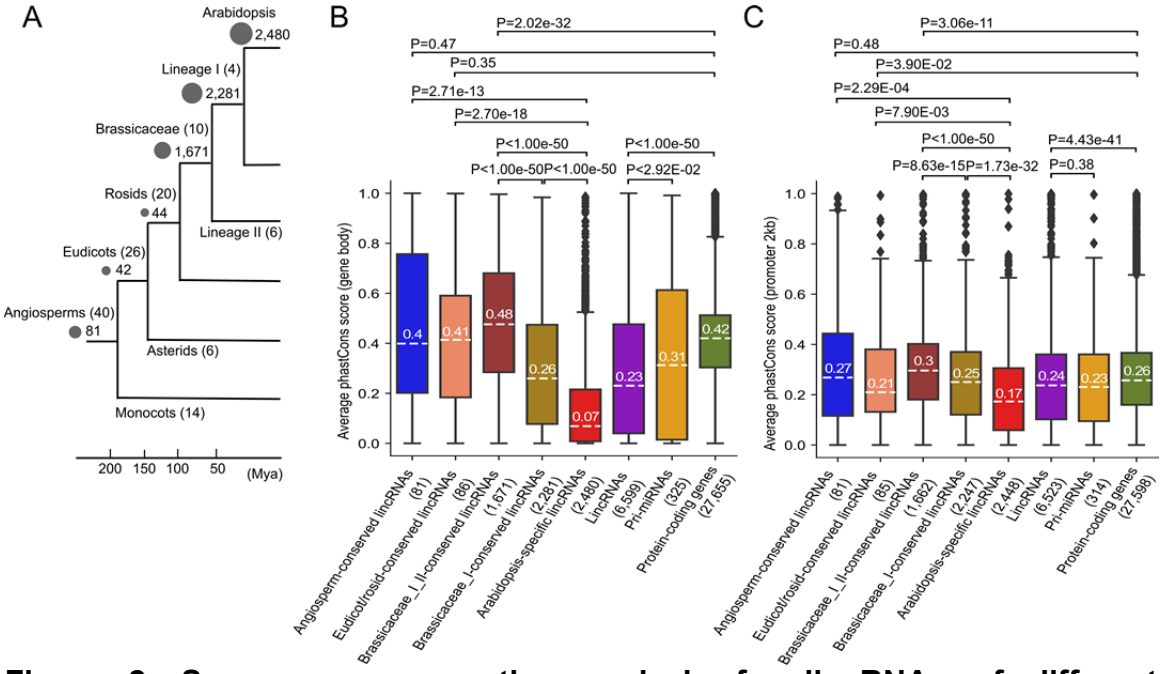

**Figure 1. Overlap and gene features of Arabidopsis lincRNA annotations. (A)** Upset plot showing the intersection of lincRNA annotation in the eight resources. Each row represents a resource, reporting in parenthesis its total number of lincRNA transcripts before merging. LincRNA annotations unique to a single resource are represented as a single circle while circles connected by lines represent the intersection of lincRNA loci shared between various resources. The bar chart indicates the number of unique lincRNA loci and intersectional lincRNA loci, displaying only intersections that contain at least ten lincRNA loci. More complex overlapping patterns are not shown. **(B)** The pie chart shows the

173 proportion of lincRNA loci supported by one or more resources. **(C)** The
174 distribution of exon number for all lincRNA transcripts (purple), protein-coding
175 transcripts (green), transcripts of lincRNAs supported by single resource (purple)
176 and multiple resources (purple). Single exon and multiple exons are shown in
177 dark and light colors, respectively. **(D)** The distribution of transcript length for
178 lincRNAs (purple) and protein-coding genes (green).
179

180 **Contrasting patterns of sequence conservation for lincRNAs belonging to**
181 **different evolutionary age categories**

182 To assess the evolutionary conservation of Arabidopsis lincRNAs within flowering
183 plants, DNA sequence similarity searches were performed by comparing our set
184 of lincRNAs with the genomes of 40 plant species (see Materials and Methods,
185 Supplemental Table S1). Among the 6,599 lincRNAs, 2,480 lincRNAs were
186 restricted to Arabidopsis and named Arabidopsis-specific lincRNAs. The other
187 lincRNAs were classified into four evolutionary age categories according to the
188 presence of homologs in closely and more distantly related species (Figure 2A).
189 We found 81 lincRNAs with at least one homolog in eudicots and in monocots
190 and therefore conserved during 180 million years (MY) of evolution (Beilstein et
191 al., 2010; Zhang et al., 2020), defined as angiosperm-conserved lincRNAs. Forty-
192 two lincRNAs were conserved in eudicots with at least one homolog in rosids and
193 asterids, but without homologs in monocots. Similarly, 44 lincRNAs were
194 identified only having homologs in rosids species, outside the Brassicaceae
195 family. As the lincRNAs conserved in eudicots and rosids showed highly similar
196 conservation patterns for gene body and promoter ($P<0.376$, Mann–Whitney *U*
197 test), we combined these genes in one category, called Eudicot/rosid-conserved
198 lincRNAs (86 genes). 1,671 Brassicaceae_I_II-conserved lincRNAs were present
199 in the common ancestor of Brassicaceae lineages I and II, without homologs
200 outside the Brassicaceae. Lastly, 2,281 lincRNAs were identified only having
201 homologs in the Brassicaceae I lineage (Brassicaceae_I-conserved lincRNAs).
202 The monotonous decrease in the number of lincRNAs for the older age

203 categories suggests there is a continuous birth of lincRNA loci at the species
204 level, with only a small fraction showing deep conservation in other plant families
205 or orders.
206




207
208 **Figure 2. Sequence conservation analysis for lincRNAs of different**
209 **evolutionary age categories. (A)** Simplified species tree reporting the number
210 of lincRNAs found for the different age categories, which are also indicated by
211 the grey circle sizes. Numbers in parenthesis report the number of genomes
212 included per clade to assess sequence similarity and define a lincRNA's age
213 category. **(B)** Sequence conservation (measured by the average phastCons
214 score over the gene body) for different lincRNA age categories and gene types
215 (lincRNAs, pri-miRNAs, and protein-coding genes). PhastCons score ranges
216 from 0 (not under selection) to 1 (strong negative selection). P-values for pairwise
217 Mann–Whitney *U* test are shown using the horizontal lines connecting the series.
218 **(C)** Boxplot showing the average phastCons score for promoter regions (2kb
219 upstream of transcription start site) of different lincRNA age categories and gene
220 types (lincRNAs, pri-miRNAs, protein-coding genes). P-values for pairwise
221 Mann–Whitney *U* test are shown using the horizontal lines connecting the series.
222

223 Apart from detecting lincRNA homologs, we also evaluated the evolutionary
224 conservation of lincRNA loci, pri-miRNAs, where globally the 21-24 miRNA
225 sequence is the only conserved region at the nucleotide level, and protein-coding
226 genes using phastCons scores (Siepel et al., 2005), which report the probability
227 for each nucleotide to evolve neutrally or under negative, or purifying, selection
228 (low or high phastCons score, respectively). In contrast to the age categories
229 which are based on finding similar homologous sequences in other plant
230 genomes, phastCons works by fitting a two-state hidden Markov model to a
231 genome-wide multiple sequence alignment and predicting, based on the pattern
232 of nucleotide substitutions, conserved elements representing sites under
233 purifying selection (Siepel et al., 2005). For lincRNA, pri-miRNA, and protein-
234 coding gene loci, we compared phastCons scores for the gene body and the
235 promoter region (2kb upstream of transcription start site) (Supplemental Data Set
236 S2). In general, for gene bodies, lincRNAs are significantly less conserved than
237 pri-miRNAs and protein-coding genes (Figure 2B, purple, yellow, and green
238 series). However, we found that Angiosperm- and Eudicot/rosid-conserved
239 lincRNAs were as conserved as protein-coding genes. Brassicaceae_I_II-
240 conserved lincRNAs showed even significantly higher conservation levels than
241 protein-coding genes, indicating high levels of purifying selection. As expected,
242 Arabidopsis-specific lincRNAs show the lowest phastCons scores.
243 In contrast to the differences in gene body, the promoter scores are comparable
244 for lincRNA, pri-miRNA, and protein-coding genes (Figure 2C, purple, yellow and
245 green series). The level of purifying selection on promoter regions is similar in
246 Angiosperm- and Eudicot/rosid-conserved lincRNAs as well as in protein-coding
247 genes. We observed that conservation levels were again significantly higher in
248 the Brassicaceae_I_II-conserved lincRNA promoters than protein-coding gene
249 promoters. Taken together, these results indicate that Brassicaceae_I_II-
250 conserved lincRNAs stand out in the level of purifying selection acting on these
251 loci, both in the gene body and in their promoter, suggesting that the primary
252 sequence of the RNA and its promoter regulation is a critical element for lincRNA
253 function.

254

255 **Large-scale transcriptome analysis reveals highly-specific lincRNA**
256 **expression in roots**

257 Increasing evidence supports the tissue or cell-type specific role of lincRNA gene
258 functions (Liu et al., 2012; Li et al., 2016; Cheng et al., 2017; Rai et al., 2019). To
259 characterize spatiotemporal Arabidopsis gene expression patterns also
260 considering lincRNAs, we curated, processed, and integrated 791 RNA-seq
261 samples to construct a genome-wide gene expression atlas covering a wide
262 range of tissues, developmental stages, and stress conditions (Supplemental
263 Table S2 and Supplemental Data Set S3). To find a threshold of detectable
264 expression above background, we normalized all data to establish detectable
265 expression levels for protein-coding genes and lincRNAs (Ramskold et al., 2009;
266 Li et al., 2016) (see Materials and Methods, Supplemental Figure S2A). A
267 normalized transcripts per million (TPM) value >= 0.2 was considered to define
268 5,586 (84.6%) expressed lincRNAs and 284 (87.4%) pri-miRNAs, whereas a
269 TPM >= 2 was used to define 26,254 (94.9%) expressed protein-coding genes.
270 Globally, the expression breadth, defined as the number of samples in which a
271 gene is expressed, was lower for lincRNAs (median: 12/791 samples) compared
272 to pri-miRNAs (median: 35.5/791 samples) and protein-coding genes (median:
273 648/791 samples) (Figure 3A). Although the identification of lincRNAs is known to
274 be impacted by the sequencing depth (Cabili et al., 2011; Liu et al., 2012; Sun et
275 al., 2017), we did not observe a clear correlation between the expression breadth
276 and sequencing depth (Pearson correlation coefficient, r=0.04, p=0.86)
277 (Supplemental Figure S2B). This result indicates the reported expression values
278 are not strongly biased by study-specific technical parameters. As previously
279 reported, the expression level of lincRNAs is generally lower than that of pri-
280 miRNAs and protein-coding genes (Figure 3B). The lincRNA loci part of the two
281 “older” categories (Angiosperm and Eudicot/rosid-conserved lincRNAs) showed
282 significantly higher expression levels compared with the three “younger”
283 categories of lincRNAs (Brassicaceae_I_II-conserved, Brassicaceae_I-conserved

284 and Arabidopsis-specific lincRNA). We also observed a six-fold difference in the
285 fraction of expressed genes depending on their annotation source: lincRNAs
286 annotated in only one resource where more frequently found as undetectable
287 compared to lincRNAs annotated in two or more resources (20% versus 3%,
288 respectively).
289 To identify groups of samples with similar expression patterns, 769 RNA-Seq
290 samples were split into 22 expression clusters using meta-data curation (22
291 samples were discarded, see Materials and Methods). Globally, these clusters
292 group Arabidopsis (Col-0) samples according to the organ or tissue considered
293 and the stress, either biotic or abiotic, applied (Supplemental Data Set S3). As
294 lincRNAs are proposed to exert their role in a tissue-specific manner (Liu et al.,
295 2015), the tissue-specificity index (tau) (Yanai et al., 2005) for each lincRNA, pri-
296 miRNA and protein-coding gene was calculated to estimate expression specificity
297 between clusters (Figure 3C). LincRNAs were more specifically expressed than
298 pri-miRNAs and protein-coding genes, with the median tau scores of 0.97, 0.95
299 and 0.74, respectively ($P<0.001$, Mann–Whitney U-test). Forty-six percent of all
300 expressed lincRNAs (2,573/5,563) with a tau score ≥ 0.97 were defined as
301 lincRNAs showing highly-specific expression (Supplemental Data Set S4). Even
302 though we did not observe a uniform distribution of these highly-specific lincRNA
303 over the 22 clusters (Figure 3D), especially cluster 19 (root), 18 (cell line light
304 induction) and 3 (seedling heat) have the largest number of highly-specific
305 lincRNAs. Between 28 and 42% of the highly-specific expressed lincRNAs were
306 present in the different evolutionary categories, with the highest fraction in the
307 Arabidopsis-specific lincRNAs (Figure 3E). Noteworthy, expression cluster 19,
308 containing 66 samples covering both whole root and specific root cell types (Li et
309 al., 2016), contains the highest fraction (14-17%) of highly-specific lincRNAs in
310 the younger age categories (Figure 3E), hinting towards a role of these lincRNAs
311 in root tissues.
312

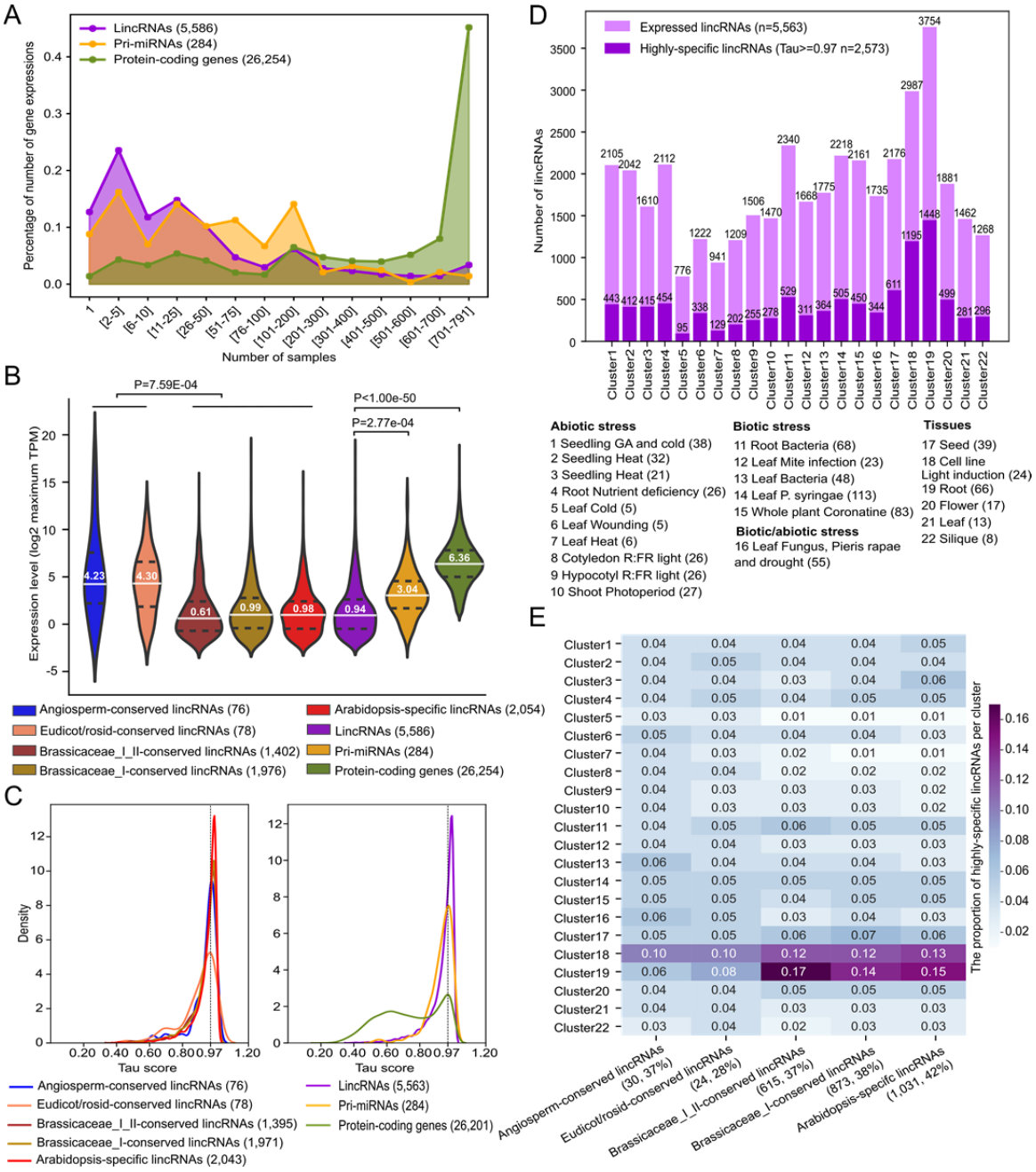



**Figure 3. Expression analysis of lincRNAs. (A)** Line chart showing the distribution of expression breadth for lincRNAs, pri-miRNAs and protein-coding genes across all samples **(B)** Distribution of the maximum TPM expression levels for different lincRNA age categories and gene types (lincRNAs, pri-miRNAs and protein-coding genes). **(C)** Distribution of tissue specificity tau scores for different lincRNA age categories and gene types. Tau scores range from zero to one, where zero means widely expressed, and one means very specifically expressed

321 (detectable in only one cluster). The black dotted line represents a tau score of
322 0.97. **(D)** Bar chart reporting the number of expressed and highly-specific
323 expressed lincRNAs for the different expression clusters. Cluster numbers and
324 descriptions are shown below the chart, with numbers in parenthesis indicating
325 the number of samples present per cluster. **(E)** Heatmap showing the proportion
326 of highly-specific expressed lincRNAs in each cluster for each lincRNA age
327 category. Cluster descriptions are the same as in panel D. Numbers in
328 parenthesis report the number of genes per age category together with the
329 fraction of lincRNAs showing highly-specific expression.
330
331 To validate these tissue-specific expression patterns, we verified the expression
332 of previously characterized Arabidopsis lincRNAs (Supplemental Table S3). The
333 lincRNA *SVALKA*, which was identified in a cold-sensitive region of the
334 Arabidopsis genome, was maximally expressed in the cold stress-related cluster
335 5 in our study (Kindgren et al., 2018). In addition, the lincRNA *MARS*, which is
336 involved in the response to abscisic acid, seed germination and root growth
337 under osmotic stress (Roule et al., 2022b), was found to be expressed at high
338 levels in root cluster 19 as well as cluster 11 (root bacteria) and cluster 17 (seed).
339 The lincRNA *ASCO*, reported to be involved in lateral root formation and
340 response to pathogens (Bardou et al., 2014; Rigo et al., 2020), was found to be
341 widely expressed, including cluster 11, which contained samples reporting
342 bacterial flagellin stress responses. The Arabidopsis lncCOBRA1, also conserved
343 in *Brassica rapa* (Kramer et al., 2022) and playing a role in seed germination,
344 was found in our set of Brassicaceae_I_II-conserved lincRNA and showed the
345 highest expression in cluster 17 (seed) and cluster 3 (seedling heat).
346 Among the highly-specific lincRNAs, the vast majority (66%) were derived from
347 one resource, with lincRNAs identified in Liu et al. 2012 being most abundant
348 (30%, Supplemental Figure S3A-B). Taken together, these findings indicate that
349 the expression clusters offer a good starting point for the context-specific
350 characterization of known and novel lincRNAs. Furthermore, young age
351 categories, including the Brassicaceae-conserved and Arabidopsis-specific

352 lincRNAs, show a bias for expression specificity in the root samples, with more
353 than half (1,448/2,573) of the highly-specific lincRNAs active in root, suggesting a
354 potentially diverse role of lincRNA networks in root growth and development.
355

356 **Experimental TF-lincRNA regulatory networks reveal active and complex**
357 **gene regulation of Arabidopsis lincRNA genes**

358 To integrate lincRNAs into epigenetic and transcriptional networks, we compared
359 the chromatin states inferred by (Liu et al., 2018; Hazarika et al., 2022) for the
360 different lincRNA gene sets delineated in our study. These chromatin states
361 report a combination of multiple epigenetic marks along the genome and offer
362 detailed insights in the locations and functions of regulatory regions and genes.
363 Globally, lincRNAs show enrichment for vastly different chromatin states
364 compared to protein-coding genes (Figure 4A). Chromatin states 33-34, typically
365 associated with DNA methylation, repressive histone modifications and
366 transposable elements, were strongly overrepresented in angiosperm-conserved
367 lincRNAs, and to a lesser extent in lincRNAs showing highly-specific expression
368 patterns. Chromatin states 31-32, also associated with DNA methylation and
369 repressive histone modifications were strongly overrepresented in Arabidopsis-
370 specific lincRNAs. Interestingly, these repressive chromatin marks were less
371 enriched in the Brassicaceae-I_II-conserved lincRNAs, where chromatin states
372 13, 19 and 20 were most strongly overrepresented. State 13 refers to Polycomb
373 group mediated deposition of trimethylation of the lysine 27 of histone H3
374 (H3K27me3) while states 19-20 denote accessible chromatin and TF ChIP
375 binding. Chromatin states 19-20, but also states 31-32 and 34, were enriched in
376 highly-specific lincRNAs. Although we found no evidence of a correlation
377 between lincRNA age categories and chromatin states, our results revealed that
378 a significant number of Brassicaceae_I_II-conserved lincRNAs have epigenetic
379 signatures associated with Polycomb regulation and TF binding in accessible
380 chromatin.

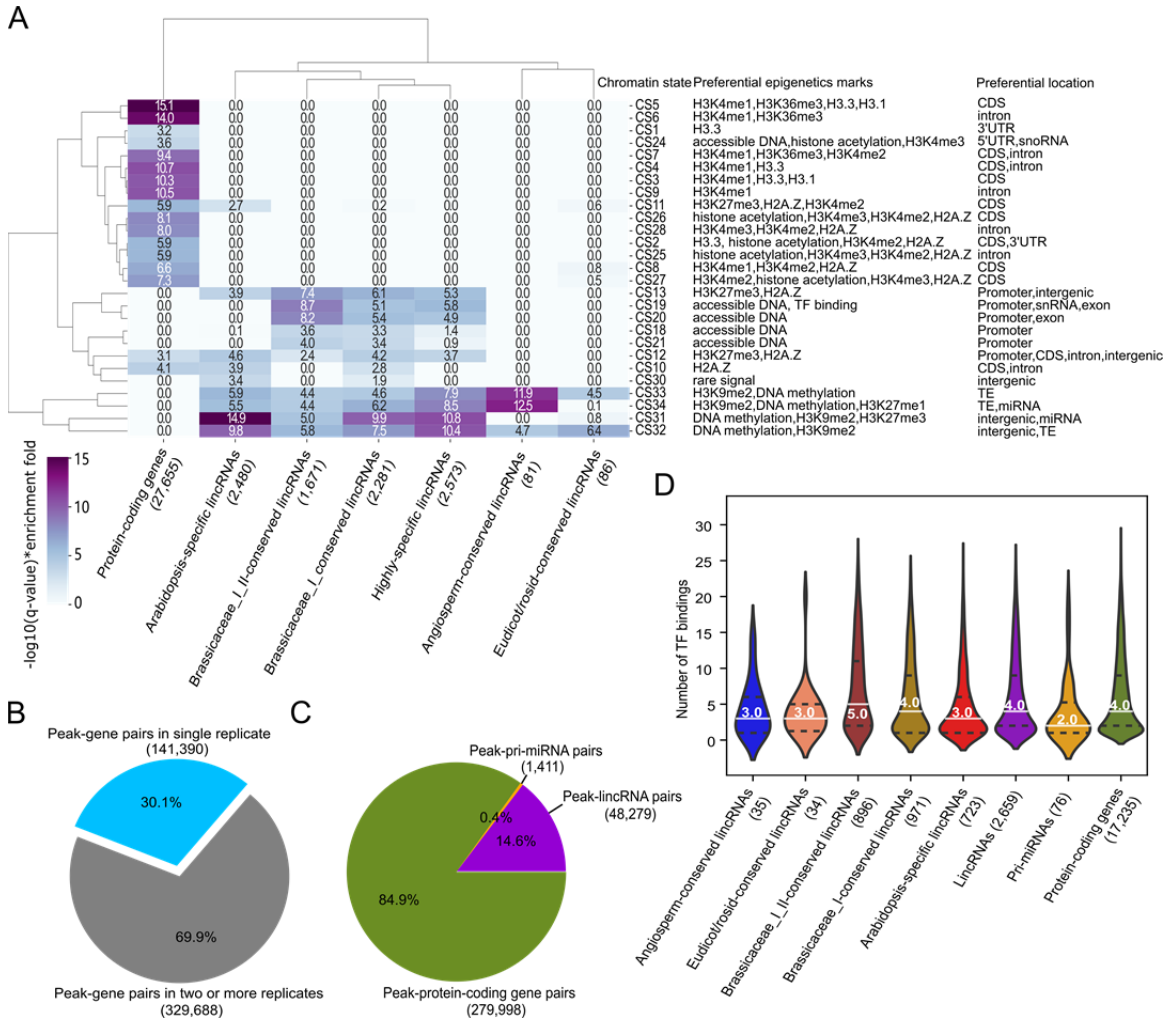



**Figure 4. Chromatin state and TF ChIP-Seq peak annotation for different gene types (A)** Dendrogram showing the enrichment for different lincRNA gene sets (x-axis) towards different chromatin states (CS) (y-axis). The values report the product of -log10(q-value) and the enrichment fold. Only significant enrichment values are reported (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 bindings for lincRNA evolutionary age categories and gene types (lincRNAs, pri-miRNAs and protein-coding genes).

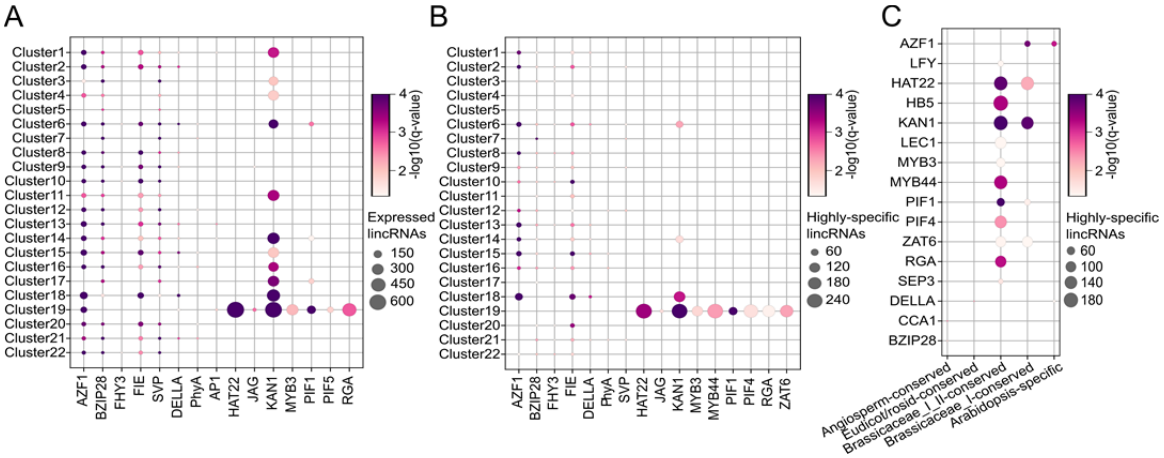
394 Based on the specific expression profiles for different lincRNA genes, the
395 chromatin state information, as well as the high levels of promoter conservation,
396 we next decided to integrate TF chromatin immunoprecipitation (ChIP) data to
397 further identify the regulators controlling lincRNAs. Before we used TF ChIP data
398 to characterize the organization, complexity, and evolution of TF binding for
399 protein-coding genes (Heyndrickx et al., 2014), hence we here re-processed
400 publicly available ChIP-Seq to identify TF binding events potentially controlling
401 lincRNA gene expression. A total of 114 TF ChIP-Seq datasets, covering 45 TFs
402 with at least two biological replicates, were reprocessed (Supplemental Table S4).
403 Starting from our genome annotation containing protein-coding genes, pri-
404 miRNAs and the high-confidence lincRNAs, ChIP-Seq peaks were assigned to
405 the closest genes. A gene was defined as a potential target gene for a profiled TF
406 if it was the closest to at least one ChIP-Seq peak of that TF (see Material and
407 Methods). Based on all 471,078 TF-target gene interactions identified from the
408 114 ChIP-Seq datasets, 70% of the peak-gene pairs were confirmed by two or
409 more replicates and were kept to construct a robust TF- target gene regulatory
410 network (Figure 4B). While most peaks were associated with protein-coding
411 genes, we identified 48,279 (14.6%) TF peak – lincRNA regulatory interactions
412 (Figure 4C). Given the strong localization bias for TF binding in Arabidopsis
413 (Heyndrickx et al., 2014; Yu et al., 2016), we only considered target genes
414 localized relative to the peak midpoint within a 2kb window (Supplemental Figure
415 S4A), retaining 95.4% of the binding events (43,233 TF-lincRNA interactions, see
416 Supplemental Data Set S5). Globally, 2,659 lincRNAs had one or more TF peaks.
417 Twenty-two out of the 45 TFs, belonging to the bHLH, HD-Zip, bZIP, C2H2,
418 GRAS, MYB, NAC, NF-YB and NF-YC TF families, were each associated with at
419 least 200 lincRNAs (Supplemental Table S5).
420 (Haudry et al., 2013) identified over 90,000 conserved non-coding sequences
421 (CNS) in Arabidopsis that show evidence of transcriptional and post-
422 transcriptional regulation. Comparing these CNS with the ChIP-Seq peaks of
423 target genes revealed that most of the TF binding events close to lincRNAs
424 (78.1%) and protein-coding genes (73.7%) overlapped with a CNS

425 (Supplemental Figure S4B). This fraction was much lower for pri-miRNAs
426 (57.0%). The highest fraction of ChIP-Seq peaks containing a CNS was detected
427 for Brassicaceae_I_II-conserved lincRNAs loci (91.8%), indicating that these TF
428 binding events are evolutionary constrained and potentially functional.
429 Considering the different gene types, the median number of TF binding events
430 per locus was higher for protein-coding genes and lincRNAs (median of 4 TFs)
431 compared to pri-miRNAs, suggesting these genes are differently controlled
432 (Figure 4D). Brassicaceae-conserved lincRNAs have more TF binding events
433 than lincRNAs in any other age categories (Figure 4D). Furthermore, we
434 observed a positive correlation between TF binding frequency and the
435 conservation of lincRNA gene body (r=0.27, P=8.11e-92) (Supplemental Figure
436 S4C). More precisely, a large faction of lincRNAs without any TF binding event
437 also has very low phastCons scores close to zero, indicating that these genes
438 are not under purifying selection. We observed a neglectable correlation between
439 the number of TF binding events and tissue specificity (r=0.02, p=1.69e-01).
440 Altogether, lincRNAs experiencing stronger levels of purifying selection are
441 bound, and potentially regulated, by more TFs, independent of their tissue
442 specificity (Supplemental Figure S4D). Taken together, we generated a multi-
443 level genome-wide characterization covering chromatin state information,
444 promoter conservation, ChIP-based TF binding and CNSs, for all detectable
445 lincRNA across >700 expression samples, permitting to rapidly define the
446 biological context and relevance of lincRNAs in Arabidopsis regulatory networks.
447

448 **MYB44, PIF4 and KAN1 regulate Arabidopsis lincRNAs in different root cell**
449 **types**

450 To identify TF-lincRNA regulatory interactions active in specific cellular contexts,
451 we used the previously defined expression clusters to combine TFs regulation
452 and lincRNA expression in specific organs, tissues, or stress conditions.
453 Considering all expressed lincRNAs and all TF peak-based regulatory
454 interactions described above, we tested if specific expression clusters are

455 overrepresented for lincRNAs controlled by specific TFs (Y-axis and X-axis in
456 Figure 5A-B, respectively). We observed a significant overrepresentation for
457 KANADI1 (KAN1) binding to lincRNA loci in 11 clusters, of which clusters 19
458 (root), 18 (cell line light induction), 14 (leaf *P. syringae*) and 6 (leaf wounding)
459 showed the most significant overlaps (Figure 5A). Comparing the different
460 clusters revealed that root cluster 19 contained numerous enriched TFs,
461 including Arabidopsis Zinc-Finger protein 1 (AZF1), JAGGED (JAG),
462 Phytochrome-Interacting Factor 5 (PIF5), Repressor of GA (RGA), PIF1, MYB3,
463 KAN1 and HAT22. The observed patterns of TF-binding in different expression
464 clusters were unique for lincRNAs and highly dissimilar compared to TF binding
465 enrichment for protein-coding genes (Supplemental Figure S5A).




466

467 **Figure 5. Overview of TF-lincRNA regulatory interactions in different**
468 **expression clusters and age categories.** The dot sizes represent the number
469 of the lincRNAs while the color represents the statistical significance. Cluster
470 descriptions are the same as in Figure 3D. **(A)** Bubble chart showing the
471 enrichment of TF binding for expressed lincRNAs in different expression clusters.
472 **(B)** Bubble chart showing the enrichment of TF binding for highly-expressed
473 lincRNAs in different expression clusters. **(C)** Bubble chart showing the
474 enrichment of TF binding for expressed lincRNAs in different age categories.

475

476 Focusing on the set of 2,573 lincRNAs showing highly-specific expression, most
477 of the TF enrichments for root cluster 19 remained (Figure 5B). PIF1, PIF4, KAN1,

478 HAT22, MYB3, MYB44, ZAT6 and RGA showed the largest overlap and all these
479 TFs, apart from MYB3 and PIF1, contained >160 lincRNA target genes (Figure
480 5B and Supplemental Table S6). We confirmed that these eight TFs were all
481 expressed in one or more samples of the root expression cluster 19
482 (Supplemental Figure S5B) and earlier studies have reported that these TFs,
483 apart from HAT22, are involved in root development in Arabidopsis (Hawker and
484 Bowman, 2004; Devaiah et al., 2007; Moubayidin et al., 2016; Tominaga-Wada
485 and Wada, 2016; Zhao et al., 2016; Li et al., 2022). Comparing TF binding for the
486 different age categories revealed that the Brassicaceae_I_II-conserved lincRNAs
487 were most strongly enriched for these TFs (Figure 5C).
488 To validate the ChIP-based regulatory interactions for the identified TFs, we used
489 reverse transcription quantitative PCR (RT-qPCR) analysis in roots of lines
490 affected in TF expression such as overexpression or T-DNA inactivation (called
491 TF perturbation lines). We used an inducible line for KAN1 (*KAN1-GR*),
492 overexpression lines for MYB4 and PIF4 (*35S::MYB4*, 35S*::PIF4*) or quadruple
493 mutant of the *pif1*, *pif2*, *pif3*, and *pif4* (*pifq*), and a knockout mutant for RGA
494 (*rga28* T-DNA line). We then selected 27 potentially regulated lincRNAs showing
495 high expression levels in the root cluster 19 and that were targeted by several of
496 the selected TFs. For example, *LincRNA5331* and *LincRNA1119* were predicted
497 to be regulated by the four TFs. We could detect deregulation for *LincRNA5331*
498 and *LincRNA1119* in *35S::MYB44* and *pifq* roots, whereas the former was also
499 deregulated in *KAN1-GR* and the latter in *35S::PIF4* (Figure 6A, Supplemental
500 Figure S6). Overall, out of the 74 inferred regulatory interactions investigated, 36
501 were confirmed, meaning that for the tested TFs, a significant deregulation of the
502 lincRNA was found in comparison to the control ("Has a peak and is DE", in
503 Table 1, Supplemental Table S7). For 23 out of the 27 tested lincRNAs, we
504 confirmed one or more regulatory interactions (Figure 6A). The precision (i.e. the
505 positive predictive value), varied between 27-65% while the average accuracy
506 (i.e. the proportion of correct predictions among all genes examined) was 59%.
507 For eight interactions the TF peak annotation to the lincRNA was unclear ("Has a
508 putative peak and is DE" in Table 1), meaning these deregulated lincRNAs might

509 also resemble confirmed interactions. Lastly, while the 10 interactions where we
510 observed deregulation in the absence of a peak could indicate false predictions,
511 they might also represent cases of indirect regulation controlled by the perturbed
512 TF, influencing the expression of the profiled lincRNAs.
513

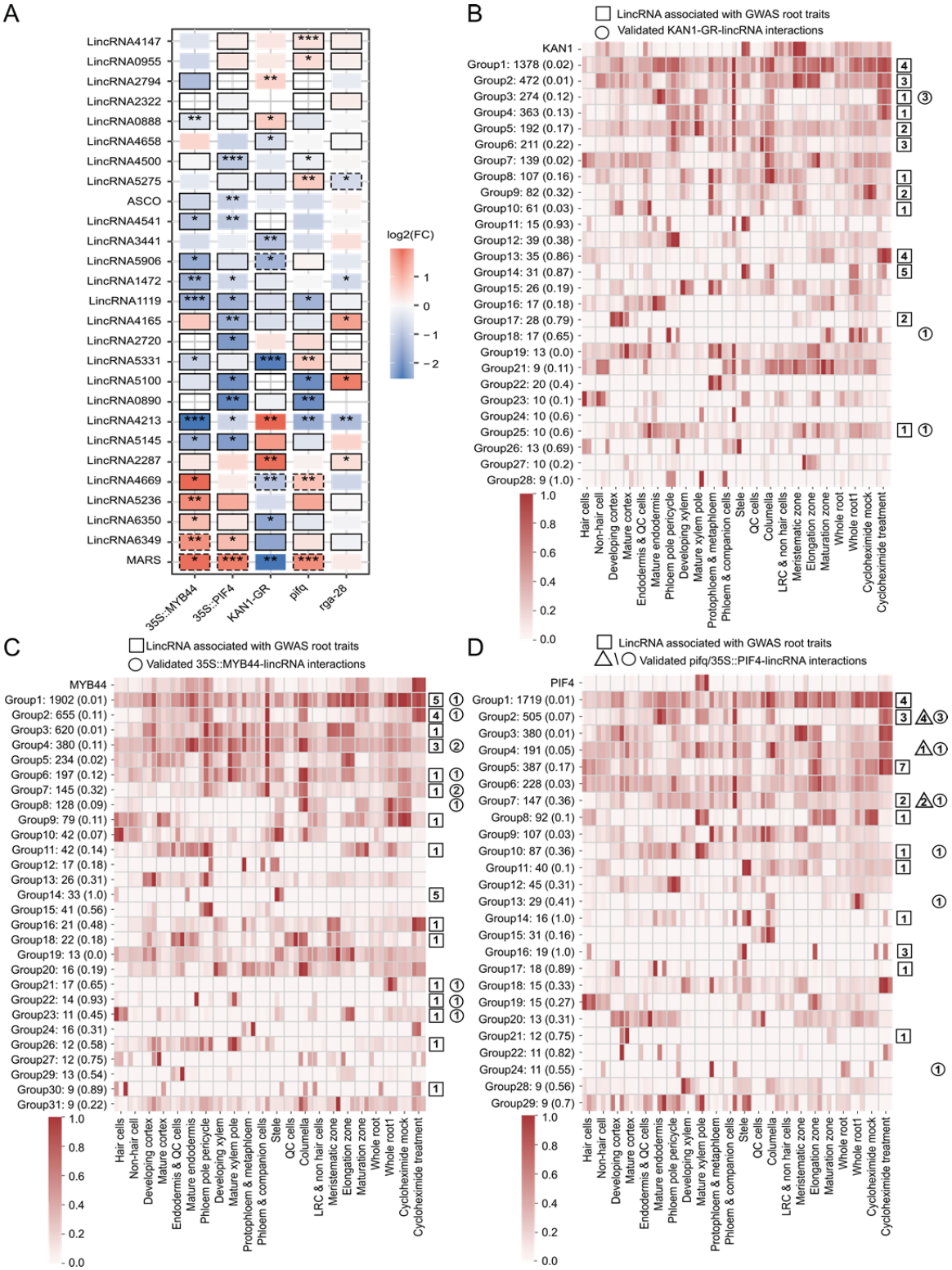



**Figure 6. Experimental validation and characterization of TF-lincRNA regulatory interactions. (A)** Heat map showing log2 fold change (FC) of lincRNA relative expression levels in transcription factor (TF) overexpressing

518 lines (35S::*MYB44*, 35S::*PIF4* and *KAN1-GR*) or TF knockout lines (*rga28* and
519 *pifq* (*pif1/pif2/pif3/pif4* quadruple mutants)) vs. control wild-type lines in 14-day
520 old roots. Expression values were determined by RT-qPCR. Asterisks indicate
521 statistically significant differences (*p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001) in an
522 unpaired two-tailed Student's t-Test (n =3). Solid boxes indicate TF ChIP peaks
523 <2kb from the lincRNA gene. Dashed boxes indicate TF ChIP peaks were
524 identified only in one ChIP-Seq replicate or the lincRNA is not the closest gene to
525 the TF peak. **(B)** Heatmap showing KAN1 ChIP-Seq peak-based target genes in
526 the cluster 19 root expression atlas. The first line shows the expression profile of
527 the TF under investigation while the other lines show groups of target genes with
528 similar expression. Next to the group number, the number of target genes is
529 reported and in parenthesis the fraction of lincRNAs per group is indicated. The
530 numbers in squares and circles report the number of GWAS root-related and RT-
531 qPCR experimentally validated lincRNA genes, respectively. **(C)** Heatmap
532 showing MYB44 ChIP-Seq peak-based target genes in the cluster 19 root
533 expression atlas. Group information is the same as in panel B. **(D)** Heatmap
534 showing PIF4 ChIP-Seq peak-based target genes in the cluster 19 root
535 expression atlas. Group information is the same as in panel B, but here triangles
536 and circles report the RT-qPCR experimentally validated lincRNA genes for the
537 *pifq* and *35S::PIF4* lines, respectively.
538
539 Based on the experimental validation results confirming that several of the root-
540 specific lincRNAs are controlled by the TFs inferred using the ChIP-based peaks,
541 we next integrated genotype-phenotype relationships from the AraGWAS catalog.
542 The aim of this analysis was to verify if root-related phenotypes have been
543 reported for regions containing lincRNAs in genome-wide association studies
544 (GWAS). After processing the significant associations from all GWAS studies
545 present in the catalog and only retaining associations overlapping with lincRNA
546 gene bodies or 2kb promoter regions (see Material and Methods), we identified
547 2,865 single nucleotide polymorphisms (SNPs) overlapping with 1,124 lincRNAs,
548 covering 147 different studies. While 61.6% of the lincRNAs had only significant

549 associations via SNPs in the promoter region, 38.4% had associations via SNPs
550 in the gene body (194 and 238 lincRNAs with significant associations only in the
551 gene body and in both gene body and promoter, respectively). After parsing and
552 summarizing phenotype information from the available trait ontology annotations,
553 we could link 23 lincRNAs to abiotic stress-related traits, 133 lincRNAs to flower-
554 related traits, 33 lincRNAs to leaf-related traits, 376 lincRNAs to root-related traits,
555 and 27 lincRNAs to seed-related traits (see Supplemental Data Set S6). The
556 reprocessed GWAS data indicate that hundreds of regions overlapping with
557 lincRNAs loci are significantly associated with different plant traits and can be
558 used to prioritize lincRNAs likely involved in specific biological processes.
559 To further support the relevance of the identified regulatory network, we focused
560 on KAN1, MYB44, and PIF4, as these regulators had several root-expressed
561 lincRNAs target genes that showed differential expression in the TF perturbation
562 lines (Table 1). Starting from the protein-coding genes and lincRNAs expressed
563 in root cluster 19 and annotated with a TF peak, we first grouped target genes
564 with similar expression in the root expression atlas. Secondly, for each group we
565 determined the fraction of lincRNAs and verified if any of these were overlapping
566 with our RT-qPCR experimentally validated lincRNAs or had root-related traits in
567 the GWAS catalog. Finally, we verified if the tissue- or cell-type specific
568 expression of the different groups was in agreement with the expression of the
569 regulatory TF. For KAN1, we identified 14 groups that fulfilled these criteria
570 (Figure 6B), of which eight contained two or more lincRNAs with root-related
571 traits and several showed strong cell-type specific expression. Group 14,
572 showing high expression in stele contained 27 lincRNAs (87%) of which five were
573 associated with GWAS root traits. The high expression of KAN1 in the
574 meristematic zone agreed with groups 1-2, 8-10, and 25, which together
575 contained 12 confirmed lincRNAs, either through root-related traits or our
576 experimental validations. For MYB44, we identified multiple groups containing
577 GWAS root-related lincRNAs showing specific expression patterns in different
578 cell types: group 4 contained multiple confirmed target genes expressed in
579 mature endodermis while group 7 contained 3 confirmed lincRNAs (one GWAS

580 root-related and two experimentally validated genes) showing high expression in
581 phloem and companion cells (Figure 6C). For PIF4, which showed highest
582 expression in mature xylem pole, groups 8 and 10 contained confirmed lincRNA
583 target genes showing similar cell-type specific expression (Figure 6D). While
584 other groups contained GWAS root-related lincRNAs showing strong expression
585 in stele (groups 11, 14 and 16), our expression data does not confirm that PIF4 is
586 actively expressed in the root stele. All GWAS root-related (58) or experimentally
587 validated (21) lincRNAs regulated by KAN1, MYB44 or PIF4 can be found in
588 Supplemental Table S8. This analysis further supports that KAN1, MYB44 and
589 PIF4 are controlling lincRNA genes showing highly specific expression in
590 different root tissues and cell types. Additionally, the overlap with genetic
591 associations hints to a role for many of these TF-controlled lincRNA loci in root
592 growth and development.
593
594
595
596 **DISCUSSION**

597 **Comprehensive annotation of Arabidopsis lincRNAs using transcriptomics**
598 **and evolutionary genomics**

599 In contrast to protein-coding genes, the characterization of lincRNAs is more
600 challenging as we lack highly curated annotations and extensive experimental
601 observations. Furthermore, the low levels of sequence conservation for the
602 majority of lincRNA loci makes it difficult to translate biological knowledge learned
603 in one species to another. Apart from co-expression network analysis, reporting
604 putative (in)direct associations between lincRNAs and other genes, information
605 about TF regulation of lincRNAs is scarce. Embedding lincRNAs in biological
606 networks has great potential to define lincRNAs linked to specific cellular or
607 morphological phenotypes.
608 Through the integration of eight lincRNA annotation resources, as well as
609 mapping various conservation, chromatin, and expression features, we presented

610 a global view on gene regulation for 5,586 expressed Arabidopsis lincRNAs. We
611 strongly focused on using replicated samples when processing high-throughput
612 datasets to obtain high-confidence gene information (Ponting and Haerty, 2022).
613 Combined with comparative genome analysis yielding information about age
614 categories and selection acting on gene bodies and promoter regions, we found
615 that different subsets of lincRNAs have distinct molecular properties. While the
616 high tissue-specificity and low levels of primary sequence conservation
617 corroborate previous findings about lincRNAs (Necsulea et al., 2014; Ma et al.,
618 2019; Palos et al., 2022), the analysis of lincRNA expression using a genome-
619 wide gene expression atlas covering nearly 800 samples revealed that lincRNA
620 expression is widespread in different organs and not restricted to stress
621 conditions (Jha et al., 2020). Focusing on lincRNAs exhibiting highly-specific
622 expression revealed that the root expression cluster contained a unique set of
623 1448 lincRNAs, which is the highest number among all expression clusters we
624 studied. When intersecting the different age categories with the expression
625 clusters, Brassicaceae_I_II-conserved lincRNAs covered a large fraction of these
626 root-expressed lincRNAs. While Palos and colleagues also identified a large
627 number of context-specific lincRNAs expressed in root tip or meristem (Palos et
628 al., 2022), our age category analysis revealed that many lincRNAs that originated
629 in the common ancestor of Brassicaceae lineages I and II, showed cell-type
630 specific expression in various root cell types.
631 We observed a clear trend of increasing levels of lincRNA gene expression when
632 going from young to old age categories. However, quantifying selection levels
633 acting on different loci revealed that the oldest age categories, showing the most
634 wide-spread and highest expression, are not undergoing the highest levels of
635 purifying selection. While older categories like angiosperm-conserved and
636 Eudicot/rosid-conserved lincRNAs had similar median phastCons scores as
637 protein-coding genes, Brassicaceae_I_II-conserved lincRNAs showed the
638 highest levels of purifying selection, both in their gene body and promoter. In
639 animals, a significant increase of dynamically expressed genes and higher levels
640 of purifying selection has been reported for older age categories (Necsulea et al.,

641 2014; Sarropoulos et al., 2019). Furthermore, massively parallel reporter assays
642 surveying thousands of human promoters revealed that tissue-specific lincRNAs
643 had fewer TF motifs compared to ubiquitously expressed genes (Mattioli et al.,
644 2019). In contrast to Arabidopsis protein-coding genes, this pattern was not found
645 for lincRNAs in our analysis (correlation between tau score and TF binding
646 frequency = 0.02). These results suggest, based on the genome-wide TF binding
647 data available in Arabidopsis, that the complexity in TF control of lincRNAs is
648 different between plants and animals and that the observed pattern of highly-
649 specific expression and purifying selection for Brassicaceae_I_II-conserved
650 lincRNAs deviates from the global trends observed in animals. Therefore, these
651 unique properties suggest that these lincRNAs, which only emerged 42 million
652 years ago, are integrated in a different manner in plant networks when compared
653 to young lincRNAs in animals.
654

655 **Multi-omics regulatory annotation of Arabidopsis lincRNAs**

656 The regulatory annotation using genome-wide chromatin state information
657 revealed that protein-coding genes and lincRNAs have distinct chromatin
658 signatures. For the latter, the enriched chromatin states were largely variable
659 between, and sometimes within, the different age categories. While states
660 associated with DNA methylation and repressive histone modifications were most
661 strongly overrepresented in the youngest (Arabidopsis-specific) and oldest
662 (angiosperm-conserved) lincRNA age categories, states denoting polycomb
663 group mediated deposition of H3K27me3 and accessible chromatin were most
664 enriched for Brassicaceae_I_II-conserved lincRNAs. H3K27me3 is a repressive
665 covalent histone modification resulting from the activity of Polycomb repressive
666 complexes. It was recently shown that a reduction in H3K27me3 levels leads to a
667 decrease in the interactions within Polycomb-associated repressive domains,
668 resulting in a global reconfiguration of chromatin architecture and transcriptional
669 reprogramming during plant development (Huang et al., 2021). Chromatin
670 accessibility is a hallmark of regulatory DNA as it allows sequence-specific

671 binding of TFs, key components of transcriptional regulatory networks (Schmitz
672 et al., 2022). The association of these chromatin states with specific sets of
673 lincRNAs strongly indicates active transcriptional regulation.
674 To identify context-specific TF regulation potentially driving the highly-specific
675 expression observed for many lincRNAs, we reprocessed, filtered and annotated
676 114 ChIP-Seq experiments covering 45 TFs, yielding a TF-lincRNA gene
677 regulatory network containing 2,659 lincRNAs and 43,233 interactions. To assess
678 the potential functionality of these inferred regulatory interactions, we overlapped
679 CNSs identified using nine Brassicaceae genomes, which confirmed that TF
680 peaks close to lincRNAs show similarly high levels of sequence constraint
681 compared to protein-coding genes (74-78%). Furthermore, TF binding events
682 close to Brassicaceae_I_II-conserved lincRNAs showed the highest levels of
683 CNS conservation (92%), which agrees with the very high phastCons promoter
684 scores we observed for this age category. While Palos and co-workers reported
685 that CNSs significantly correlated with gene bodies of Brassicaceae-conserved
686 lincRNAs (Palos et al., 2022), our results revealed that also lincRNA promoters
687 and TF binding sites are strongly conserved for Brassicaceae_I-II-conserved
688 lincRNAs. Ultraconserved CNSs, frequently associating with TF binding sites for
689 key plant regulators controlling essential biological processes, have been
690 identified for thousands of protein-coding genes (Van de Velde et al., 2016). Our
691 results suggest that such deep conservation of cis-regulatory elements is
692 extremely rare for lincRNAs, as only a small number of lincRNA genes show
693 deep evolutionary conservation. Such deeply conserved CNSs for protein-coding
694 genes frequently occur in divergent gene pairs, where they form mini-regulons
695 representing conserved transcriptional units of co-regulated and co-expressed
696 neighboring genes (Van de Velde et al., 2016). It is currently unclear if such
697 conserved transcriptional regulons also exist for lincRNA loci and could explain
698 the observed patterns of positionally-conserved but sequence-diverged lncRNAs
699 (Mohammadin et al., 2015).
700 The integrated TF binding information showed that promotors of lincRNA differ
701 strongly from those of protein-coding genes, but also revealed high levels of

heterogeneity among the different age categories. In animals, promoters of protein-coding genes contain more TF binding sites than those of lincRNAs, suggesting a stronger and more complex transcriptional regulation of the former (Necsulea et al., 2014; Sarropoulos et al., 2019). When comparing the number of TF binding events for the different gene types in Arabidopsis, we observed no difference in TF binding frequency for protein-coding genes and lincRNAs, indicating that lincRNAs are also regulated in a complex manner in plants. While no correlation between lincRNA expression tissue-specificity and TF binding frequency was found, a positive correlation between TF binding frequency and the level of purifying selection was observed. Again, Brassicaceae_I_II-conserved lincRNAs stood out having the highest number of binding TFs, suggesting that neither the age nor the expression of a lincRNA, but its importance for plant fitness is a major factor in determining its regulatory complexity. While our results confirm that broadly expressed protein-coding genes, showing high expression breadth, are positively correlated with the number of regulating TFs (Heyndrickx et al., 2014), we did not observe this global trend for lincRNAs, indicating that the regulatory properties of TF control for plant protein-coding genes and lincRNAs are different. Although the number of Arabidopsis TFs profiled using ChIP-Seq may be considered as limited, future research will have to address whether the complexity of TF control varies for lincRNAs, as well as protein-coding genes, active in different organs, tissues, or stress conditions.

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

Through integration of our spatiotemporal expression clusters and the TF-lincRNA gene regulatory network, we identified eight TF regulators showing a significant enrichment for TF binding close to lincRNAs specifically expressed in roots. While the overlap between TF ChIP-Seq peaks and CNSs gave an indirect indication of the potential functionality of TF binding sites close to hundreds of

732 lincRNA loci, we experimentally validated a set of inferred regulatory interactions,
733 focusing on 27 root-expressed lincRNAs and 5 TF perturbation lines. The number
734 of confirmed regulatory interactions as well as the positive prediction values
735 found for the tested TFs and lincRNAs here (27-65%) are 3 to 4 times higher
736 than the fraction of TF-bound protein-coding genes also showing deregulation
737 reported for a set of TF regulators involved in flowering (7-22%) (O'Maoileidigh et
738 al., 2014). Compared to the discovery rates obtained for large-scale phenotypic
739 screens of insertional lines (1.3%) (Ransbotyn et al., 2015), our discovery rates
740 for deregulation are 20-50 higher. Globally, for 85% of the profiled lincRNAs one
741 or more confirmed regulatory interactions were found. These findings indicate
742 that the ChIP peak-based inference of TF regulation is a promising approach to
743 characterize TF regulation of lincRNAs. Additional deregulated lincRNAs lacking
744 a TF peak were also identified, which might be due to an indirect effect caused
745 by crosstalk between different regulators in the Arabidopsis root as well as the
746 type of mutations chosen in the perturbed TF lines (e.g. partners lacking in
747 overexpressing lines, compensatory effects in gene families) (Heyndrickx et al.,
748 2014).
749 While detailed functional characterizations of lincRNA genes are scarce, the
750 integration of GWAS information allowed us to identify genetic associations for
751 1,124 lincRNAs covering various traits. While this number, corresponding to a
752 frequency of 9%, is slightly lower compared to that of protein-coding genes
753 associated with a specific trait in the AraGWAS catalog (3,030/27,655 = 11%), it
754 does confirm the great potential of this largely untapped resource to biologically
755 characterize lincRNAs potentially controlling different plant traits. As shown for
756 KAN1, MYB44 and PIF4, multiple groups of co-expressed lincRNAs were
757 identified bound by one of these TFs. Most of these groups showed strong cell-
758 type specific expression and contained lincRNAs that were annotated with root-
759 related traits.
760 Taken together, the integration of different gene annotations combined with
761 information about evolutionary conservation, selection, expression, TF regulation
762 and GWAS data yielded new insights on the biological relevance of hundreds of

763 Arabidopsis lincRNAs and offers a promising strategy to identify lincRNAs
764 involved in different aspects of plant biology.
765
766

767 **MATERIALS AND METHODS**

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

769 Paired-end RNA-seq datasets with high sequencing depth conducted in previous
770 projects in the group at the Institute of Plant Sciences Paris-Saclay were used to
771 predicte additional lincRNAs. All data came from experiments carried out in the
772 Col0 ecotypes and involved nsra/b mutant in response to NPA/NAA treatment
773 GSE65717 and GSE116923 (Tran Vdu et al., 2016; Bazin et al., 2018)
774 modification of expression of the *ASCO* lincRNA GSE135376 (Rigo et al., 2020),
775 root tip submitted to a short phosphate starvation kinetic GSE128250 (Blein et al.,
776 2020) and a lateral root initiation kinetic from a binding essay of five time points
777 without replicates (6h, 12h, 24h, 36h and 48h after binding, unpublished). All
778 reads were quality trimmed using Trimmomatic. For each library independently,
779 cleaned reads were mapped on TAIR9 genome sequence with STAR (version
780 2.7.2a) using Araport11 as a guided annotation with the following additional
781 parameters: --alignIntronMin 20 --alignIntronMax 3000. For each alignment file,
782 StringTie (version 2.1.4) was used to predict transcripts using Araport11
783 annotations as a guide (additional parameters: -c 2.5 -j 10). GFFcompare
784 (v0.12.6) was then used to isolate the new transcripts in comparison to Araport11
785 gene annotation (retaining only class code k, m and j). The different transcripts
786 prediction were then combined using StringTie in merge (additional parameters: -
787 F 0 -T 0 -c 0 -f 0 -g 0 -i). The final set of transcripts were compared against
788 Araport11 annotation with GFFcompare removing all transcripts with a class code
789 of =, c, e, s or m. Coding potential was then assets using COME (Hu et al., 2017),
790 Coding Potential Calculator CPC (v0.9-r2), CPC2 (Kang et al., 2017) and infernal
791 (v1.1.2) (Nawrocki and Eddy, 2013) against Rfam v14.1 (Kalvari et al., 2021) with
792 default parameters. Non-coding transcripts were the ones predicted by CPC,

CPC2 and COME as non-coding and having no hits against tRNA, rRNA, snRNA or snoRNA genes in Rfam.

**Integration of lincRNA gene annotations**

*Arabidopsis* lincRNAs were collected from public databases: Araport11 (Cheng et al., 2017), CANTATAdb (Szczesniak et al., 2016), NONCODEv5 (Fang et al., 2018), PLNlncRbase (Xuan et al., 2015), three publications(Liu et al., 2012; Nelson et al., 2017; Zhao et al., 2018), and the in-house data.

The pipeline used for the identification of putative lincRNA is described in Supplemental Figure S1A. (1) LincRNA transcripts with a length of at least 200 bp were retained. (2) Only transcripts that were at least 500 bp away from any protein-coding gene were retained and considered as intergenic (Liu et al., 2012; Yamada, 2017). (3) Transcripts lacking strand information were discarded. (4) The coding potential of transcripts was assessed using the Coding-Non-Coding Index (CNCI; Version 2) (Sun et al., 2013), CPC2, and Pfam-scan (PFAM) (Finn et al., 2016), and only those transcripts fulfilling the CPC2 (cutoff < 0), CNCI (cutoff < 0) and PFAM (E-value 1e-5) criteria were retained. (6) All candidate intergenic transcripts were assigned to lincRNA loci using the GFFcompare program (Pertea and Pertea, 2020). The number of transcripts retained after each filtering step is reported in Supplemental Figure S1B.

**Evolutionary conservation analysis**

Our set of Arabidopsis lincRNAs was classified into distinct evolutionary age categories based on sequence similarity. The sequences of 6599 lincRNA loci were extracted using BEDTools getfasta v2.30.0 (Quinlan and Hall, 2010). Genome sequence data for 40 plant species were obtained from PLAZA 5.0 Dicots and PLAZA 5.0 Monocot (Van Bel et al., 2021), representing 26 eudicots (20 rosids and 6 asterids ) and 14 monocots. The 20 rosids contain 10 Brassicaceae species. LincRNA homologs were identified using sequence similarity searches against these 40 genomes using BLASTn (Altschul et al., 1990) and applying an E-value cutoff of 1e–10 (Nelson et al., 2017).

822 Classification rules were defined to construct five evolutionary age categories. A
823 lincRNA was deemed conserved in the Angiosperm evolutionary age category
824 when at least one homolog was found in eudicots and one homolog in monocots.
825 LincRNA was defined as a eudicot/rosid-conserved lincRNA with at least one
826 homolog in rosids, one homolog in asterids, and no homolog in monocots, in
827 addition with one homolog only in rosids. The lincRNA was assigned as a
828 Brassicaceae_I_II-conserved lincRNA with at least one homolog in Brassicaceae
829 lineage I, one homolog in Brassicaceae lineage II, and no homologs outside the
830 Brassicaceae species. A lincRNA that had at least one homolog in a
831 Brassicaceae lineage I species, apart from Arabidopsis, was defined as
832 Brassicaceae_I-conserved lincRNA. The last category, defined as Arabidopsis-
833 specific lincRNA, were restricted to Arabidopsis lincRNAs without homologs in
834 any of the other species.
835 We downloaded the GFF annotation file for 27,655 protein-coding genes and 325
836 pri-miRNAs from the Araport11 genome release (Cheng et al., 2017) and
837 extracted gene body sequences using BEDTools getfasta v2.30.0. We also
838 extracted the promoter region of 2kb upstream of the transcription start site for
839 lincRNAs, pri-miRNAs, and protein-coding genes. Noted that the promoter of a
840 gene was shortened when it overlapped with a nearby coding sequence. The
841 sequence conservation of gene body and promoter region per gene type was
842 evaluated by average phastCons scores, which were calculated using the
843 alignments of 20 angiosperm plant genomes (Hupalo and Kern, 2013). Genes
844 without phastCons score, for example due to missing genome alignment
845 information, were assigned a score of zero.

846 **Expression analysis of lincRNAs**

847 To generate an expression atlas for lincRNAs, pri-miRNAs and protein-coding
848 genes, we used Curse (Vaneechoutte and Vandepoele, 2019) to search and
849 curate relevant RNA-seq experiments. Details of the 18 RNA-seq experiments
850 across all 791 samples are shown in Supplemental Table S2 and Supplemental
851 Data Set S3. Next, we imported the expression metadata to the Prose tool

852 (Vaneechoutte and Vandepoele, 2019), which downloads the raw data from SRA
853 by using the SRA toolkit, performs quality control and adapter detection,
854 trimmomatic to perform adapter clipping and quality trimming and finally Kallisto
855 (Pertea et al., 2015) for quantifying transcript expression to normalized
856 transcripts per million (TPM) values. We downloaded the transcript FASTA file for
857 27,655 protein-coding genes and 325 pri-miRNAs from the Araport11 genome
858 release (Cheng et al., 2017) and retrieved transcript sequences for 6,599
859 lincRNAs using gffread (Pertea and Pertea, 2020). Gene-level atlases were
860 created by summing the TPM values of all transcripts. We retained TPM value
861 per gene across the biological replicates and took an average of TPM values per
862 gene from technical replicates, resulting in an expression atlas covering 791
863 samples.
864 We used a simulation experiment (Ramskold et al., 2009; Li et al., 2016) to
865 determine thresholds for detectable expression in TPM for protein-coding genes
866 and lincRNAs. The protein-coding genes were used as true positives and the
867 lincRNAs were used as true negatives. We calculated the false-positive rate and
868 false-negative rate at different TPM thresholds for each sample. The threshold
869 TPM of 0.5 balances the numbers of false positives and false negatives.
870 LincRNAs and pri-miRNAs with TPM >= 0.2 in at least one sample were
871 considered as expressed. A more stringent threshold TPM >= 2 was used to
872 define the expression of protein-coding genes.
873 To identify the clusters containing samples with related gene expression
874 information, a $\log_2$-transformed gene expression matrix (TPM+1) was used for t-
875 SNE clustering with perplexity 30 and n_iter = 1000 (Pedregosa et al., 2011). 22
876 clusters were retained, the largest of which contained 113 samples and the
877 smallest contained 5 samples. Twenty-two samples were not clustered and
878 removed (Supplemental Data Set S3). Expression specificity per gene type, for
879 the gene expression matrix described above, was measured using tissue-
880 specificity index (tau), defined by the following equation (Yanai et al., 2005):

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

881 Where $x_i$ was defined as the average TPM value of per cluster and $n$
882 corresponds to the number of clusters analyzed.
883 Groups of genes showing similar expression profiles reported in Figure 6 B-D
884 were identified using CAST clustering and only clusters with at least 10 genes
885 were retained (Vandepoele et al., 2009).

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

887 The enrichment of the lincRNA sets in different chromatin states (CSs) of
888 Arabidopsis from (Liu et al., 2018) was done by shuffling the lincRNA genome
889 coordinates 1000 times over the whole Arabidopsis genome. Then, we compared
890 the distribution of the number of lincRNAs expected to overlap by chance with
891 each CS with the real number of overlaps, and we used these values to calculate
892 enrichment statistics: p-value as the number of times the real overlap was higher
893 than the overlap with any of the 1000 shuffled lincRNA sets, and enrichment fold
894 as the real overlap divided by the median of overlap expected by chance (median
895 of the 1000 shuffling events). The p-value was adjusted for multiple testing using
896 Benjamini-Hochberg correction (significance level 0.05). For visualization
897 purposes, the two enrichment metrics were combined into the π-value, which is
898 the -log10(p-value)*enrichment fold.
899 TF ChIP-Seq peak coordinates were retrieved from the PlantPAN 3.0 database
900 (Chow et al., 2019) and (Song et al., 2016). The original ChIP-Seq of KAN1,
901 MYB44 and PIF4 were derived from these studies (Merelo et al., 2013; Pfeiffer et
902 al., 2014; Song et al., 2016). For each peak the closest gene was identified and
903 only peaks confirmed in two or more replicates and within a 2kb window of the
904 gene body were retained. Starting from all TF – target gene pairs (either a
905 protein-coding gene or a lincRNAs; Supplemental Data Set S5), enrichment
906 analysis was performed to identify enriched TFs in different expression clusters
907 or age categories. For all enrichment analyses the hypergeometric distribution
908 was applied and the q-value of enrichment was determined using the Benjamini–
909 Hochberg correction for multiple hypotheses testing.

910 **Identification of GWAS-associated genes**

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

919 **Validation of TF-lincRNA regulatory interactions using reverse**
920 **transcription-quantitative PCR**

921 Total RNA from five to eight 14-day roots grown vertically on solid MS ½ media
922 was extracted using TRI Reagent (Sigma-Aldrich) and digested with RNase-free
923 DNase (Fermentas) following the manufacturer's recommendations. cDNA was
924 synthetized using Maxima Reverse Transcriptase (Thermo Scientific). Expression
925 analysis by RT-qPCR was performed using SYBR Green master I (Roche) and
926 the LightCycler® 96 system following a standard protocol (40 cycles, 60°C
927 annealing). Data were analyzed using the ΔΔCt method with PP2A (PROTEIN
928 PHOSPHATASE 2A SUBUNIT A3 (AT1G13320)) as reference transcript for
929 normalization of RT-qPCR data. WT Col-0 plants grown at the same time were
930 used as sample reference. For the dexamethasone inducible KAN1-GR
931 expression lines analysis, after 14 days on MS media, the plants were transferred
932 for one day either on dexamethasone containing plates (10µM) or only DMSO
933 (dexamethasone solvent) as control. Primers used are listed in Supplemental
934 Table S9. Three biological replicates were performed per condition. Statistical
935 analyses were performed using the unpaired two-tailed Student's t-Test
936 (GraphPad prism).
937

938 **SUPPLEMENTAL DATA**

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

940 **Supplemental Figure S2.** Definition of expressed genes and influence of
941 sequencing depth.

942 **Supplemental Figure S3.** Resource annotation for highly-specific expressed
943 lincRNAs.

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

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

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

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

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

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

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

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

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

957 **Supplemental Table S7.** Overview of lincRNA peak annotation and RT-qPCR
958 validation.

959 **Supplemental Table S8.** RT-qPCR confirmed/associated with GWAS root traits
960 lincRNAs expression information for the different samples of expression cluster
961 19 (root) (corresponds to Figure 6B-D).

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

963 **Supplemental Data Set S1.** Full list of putative lincRNAs integrated from
964 different resources.

965 **Supplemental Data Set S2.** PhastCons Scores for lincRNA loci and evolutionary
966 age category.

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

969 **Supplemental Data Set S4.** The list of the highly-specific lincRNAs.
970 **Supplemental Data Set S5.** The table of TF-lincRNA interactions.
971 **Supplemental Data Set S6.** The lincRNA loci containing significant GWAS hits.
972
973 L.L., M.C., T.B., and K.V., designed the research. L.L., T.D., and T.B., performed
974 data analysis. N.M.P. performed chromatin state analysis. M.H. performed
975 experimental work. L.L., T.B., and K.V., wrote the manuscript.
976

977 **ACKNOWLEDGEMENTS**

978 We thank Mariek Dubois, Andrés Ritter, Tereza Vavrdova, Freya De Winter,
979 Rebecca De Clercq for kindly providing the Arabidopsis lines. We thank Olivier
980 Martin and Federico Ariel for providing comments on the manuscript. The IPS2 is
981 benefited from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007).
982 L.L. is supported by the China Scholarship Council for a PhD fellowship
983 (201808530499). M.H. is supported by a PhD fellowship from the Fondation pour
984 la Recherche Médicale (ECO202106013730).
985

986 **FIGURE LEGENDS: see inline figure captions**

987

988 **TABLES**

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

990

| Line | Total number of validated lincRNAs (1) | Has a peak and is DE (2) | No peak and not DE | Has a peak and is not DE | No peak and is DE | Has a putative peak and is DE (3) | Accuracy (4) | Precision | Recall |
|---|---|---|---|---|---|---|---|---|---|
| 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 |

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

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

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

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

1000

1001 **REFERENCES**


1002 **Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J.** (1990). Basic local
1003 alignment search tool. J Mol Biol **215,** 403-410.
1004 **Ariel, F., Lucero, L., Christ, A., Mammarella, M.F., Jegu, T., Veluchamy, A.,**
1005 **Mariappan, K., Latrasse, D., Blein, T., Liu, C., Benhamed, M., and Crespi, M.**
1006 (2020). R-Loop Mediated trans Action of the APOLO Long Noncoding RNA. Mol
1007 Cell **77,** 1055-1065 e1054.
1008 **Bardou, F., Ariel, F., Simpson, C.G., Romero-Barrios, N., Laporte, P., Balzergue, S.,**
1009 **Brown, J.W., and Crespi, M.** (2014). Long noncoding RNA modulates
1010 alternative splicing regulators in Arabidopsis. Dev Cell **30,** 166-176.
1011 **Bazin, J., Romero, N., Rigo, R., Charon, C., Blein, T., Ariel, F., and Crespi, M.** (2018).
1012 Nuclear Speckle RNA Binding Proteins Remodel Alternative Splicing and the
1013 Non-coding Arabidopsis Transcriptome to Regulate a Cross-Talk Between Auxin
1014 and Immune Responses. Frontiers in Plant Science **9**.
1015 **Beilstein, M.A., Nagalingum, N.S., Clements, M.D., Manchester, S.R., and Mathews,**
1016 **S.** (2010). Dated molecular phylogenies indicate a Miocene origin for Arabidopsis
1017 thaliana. Proc Natl Acad Sci U S A **107,** 18724-18728.
1018 **Bhogireddy, S., Mangrauthia, S.K., Kumar, R., Pandey, A.K., Singh, S., Jain, A.,**
1019 **Budak, H., Varshney, R.K., and Kudapa, H.** (2021). Regulatory non-coding
1020 RNAs: a new frontier in regulation of plant biology. Functional & Integrative
1021 Genomics **21,** 313-330.
1022 **Blein, T., Balzergue, C., Roule, T., Gabriel, M., Scalisi, L., Francois, T., Sorin, C.,**
1023 **Christ, A., Godon, C., Delannoy, E., Martin-Magniette, M.L., Nussaume, L.,**
1024 **Hartmann, C., Gautheret, D., Desnos, T., and Crespi, M.** (2020). Landscape of
1025 the Noncoding Transcriptome Response of Two Arabidopsis Ecotypes to
1026 Phosphate Starvation. Plant Physiol **183,** 1058-1072.
1027 **Bu, D.C., Luo, H.T., Jiao, F., Fang, S.S., Tan, C.F., Liu, Z.Y., and Zhao, Y.** (2015).
1028 Evolutionary annotation of conserved long non-coding RNAs in major mammalian
1029 species. Science China-Life Sciences **58,** 787-798.
1030 **Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn,**
1031 **J.L.** (2011). Integrative annotation of human large intergenic noncoding RNAs
1032 reveals global properties and specific subclasses. Genes Dev **25,** 1915-1927.
1033 **Chen, L., Zhu, Q.H., and Kaufmann, K.** (2020). Long non-coding RNAs in plants:
1034 emerging modulators of gene activity in development and stress responses.
1035 Planta **252,** 92.
1036 **Chen, Q.S., Liu, K., Yu, R., Zhou, B.L., Huang, P.P., Cao, Z.X., Zhou, Y.Q., and**
1037 **Wang, J.H.** (2021). From "Dark Matter" to "Star": Insight Into the Regulation
1038 Mechanisms of Plant Functional Long Non-Coding RNAs. Frontiers in Plant
1039 Science **12**.
1040 **Cheng, C.Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S., and**
1041 **Town, C.D.** (2017). Araport11: a complete reannotation of the Arabidopsis
1042 thaliana reference genome. Plant J **89,** 789-804.
1043 **Chow, C.N., Lee, T.Y., Hung, Y.C., Li, G.Z., Tseng, K.C., Liu, Y.H., Kuo, P.L., Zheng,**
1044 **H.Q., and Chang, W.C.** (2019). PlantPAN3.0: a new and updated resource for
1045 reconstructing transcriptional regulatory networks from ChIP-seq experiments in
1046 plants. Nucleic Acids Research **47,** D1155-D1163.
1047 **Deng, F., Zhang, X., Wang, W., Yuan, R., and Shen, F.** (2018). Identification of
1048 Gossypium hirsutum long non-coding RNAs (lncRNAs) under salt stress. BMC
1049 Plant Biol **18,** 23.

1050 **Devaiah, B.N., Nagarajan, V.K., and Raghothama, K.G.** (2007). Phosphate
1051 homeostasis and root development in Arabidopsis are synchronized by the zinc
1052 finger transcription factor ZAT6. Plant Physiol **145,** 147-159.
1053 **Fang, S.S., Zhang, L.L., Guo, J.C., Niu, Y.W., Wu, Y., Li, H., Zhao, L.H., Li, X.Y.,**
1054 **Teng, X.Y., Sun, X.H., Sun, L., Zhang, M.Q., Chen, R.S., and Zhao, Y.** (2018).
1055 NONCODEV5: a comprehensive annotation database for long non-coding RNAs.
1056 Nucleic Acids Research **46,** D308-D314.
1057 **Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter,**
1058 **S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J.,**
1059 **and Bateman, A.** (2016). The Pfam protein families database: towards a more
1060 sustainable future. Nucleic Acids Res **44,** D279-285.
1061 **Fukuda, M., Nishida, S., Kakei, Y., Shimada, Y., and Fujiwara, T.** (2019). Genome-
1062 Wide Analysis of Long Intergenic Noncoding RNAs Responding to Low-Nutrient
1063 Conditions in Arabidopsis thaliana: Possible Involvement of Trans-Acting siRNA3
1064 in Response to Low Nitrogen. Plant Cell Physiol **60,** 1961-1973.
1065 **Haudry, A., Platts, A.E., Vello, E., Hoen, D.R., Leclercq, M., Williamson, R.J.,**
1066 **Forczek, E., Joly-Lopez, Z., Steffen, J.G., Hazzouri, K.M., Dewar, K.,**
1067 **Stinchcombe, J.R., Schoen, D.J., Wang, X., Schmutz, J., Town, C.D., Edger,**
1068 **P.P., Pires, J.C., Schumaker, K.S., Jarvis, D.E., Mandakova, T., Lysak, M.A.,**
1069 **van den Bergh, E., Schranz, M.E., Harrison, P.M., Moses, A.M., Bureau, T.E.,**
1070 **Wright, S.I., and Blanchette, M.** (2013). An atlas of over 90,000 conserved
1071 noncoding sequences provides insight into crucifer regulatory regions. Nat Genet
1072 **45,** 891-898.
1073 **Hawker, N.P., and Bowman, J.L.** (2004). Roles for Class III HD-Zip and KANADI genes
1074 in Arabidopsis root development. Plant Physiol **135,** 2261-2270.
1075 **Hazarika, R.R., Serra, M., Zhang, Z., Zhang, Y., Schmitz, R.J., and Johannes, F.**
1076 (2022). Molecular properties of epimutation hotspots. Nat Plants **8,** 146-156.
1077 **He, H., Zhou, Y.F., Yang, Y.W., Zhang, Z., Lei, M.Q., Feng, Y.Z., Zhang, Y.C., Chen,**
1078 **Y.Q., Lian, J.P., and Yu, Y.** (2021). Genome-Wide Analysis Identified a Set of
1079 Conserved lncRNAs Associated with Domestication-Related Traits in Rice.
1080 International Journal of Molecular Sciences **22**.
1081 **Heo, J.B., and Sung, S.** (2011). Vernalization-mediated epigenetic silencing by a long
1082 intronic noncoding RNA. Science **331,** 76-79.
1083 **Heyndrickx, K.S., Van de Velde, J., Wang, C., Weigel, D., and Vandepoele, K.** (2014).
1084 A functional and evolutionary perspective on transcription factor binding in
1085 Arabidopsis thaliana. Plant Cell **26,** 3894-3910.
1086 **Hu, L., Xu, Z., Hu, B., and Lu, Z.J.** (2017). COME: a robust coding potential calculation
1087 tool for lncRNA identification and characterization based on multiple features.
1088 Nucleic Acids Res **45,** e2.
1089 **Huang, Y., Sicar, S., Ramirez-Prado, J.S., Manza-Mianza, D., Antunez-Sanchez, J.,**
1090 **Brik-Chaouche, R., Rodriguez-Granados, N.Y., An, J., Bergounioux, C.,**
1091 **Mahfouz, M.M., Hirt, H., Crespi, M., Concia, L., Barneche, F., Amiard, S.,**
1092 **Probst, A.V., Gutierrez-Marcos, J., Ariel, F., Raynaud, C., Latrasse, D., and**
1093 **Benhamed, M.** (2021). Polycomb-dependent differential chromatin
1094 compartmentalization determines gene coregulation in Arabidopsis. Genome
1095 Research **31,** 1230-+.
1096 **Hupalo, D., and Kern, A.D.** (2013). Conservation and functional element discovery in 20
1097 angiosperm plant genomes. Mol Biol Evol **30,** 1729-1744.

1098 **Jha, U.C., Nayyar, H., Jha, R., Khurshid, M., Zhou, M., Mantri, N., and Siddique,**
1099 **K.H.M.** (2020). Long non-coding RNAs: emerging players regulating plant abiotic
1100 stress response and adaptation. BMC Plant Biol **20,** 466.
1101 **Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B.** (2006). MicroRNAs and their
1102 regulatory roles in plants. Annual Review of Plant Biology **57,** 19-53.
1103 **Kalvari, I., Nawrocki, E.P., Ontiveros-Palacios, N., Argasinska, J., Lamkiewicz, K.,**
1104 **Marz, M., Griffiths-Jones, S., Toffano-Nioche, C., Gautheret, D., Weinberg, Z.,**
1105 **Rivas, E., Eddy, S.R., Finn, R.D., Bateman, A., and Petrov, A.I.** (2021). Rfam
1106 14: expanded coverage of metagenomic, viral and microRNA families. Nucleic
1107 Acids Res **49,** D192-D200.
1108 **Kang, Y.J., Yang, D.C., Kong, L., Hou, M., Meng, Y.Q., Wei, L., and Gao, G.** (2017).
1109 CPC2: a fast and accurate coding potential calculator based on sequence
1110 intrinsic features. Nucleic Acids Res **45,** W12-W16.
1111 **Ke, L.L., Zhou, Z.W., Xu, X.W., Wang, X., Liu, Y.L., Xu, Y.T., Huang, Y., Wang, S.T.,**
1112 **Deng, X.X., Chen, L.L., and Xu, Q.** (2019). Evolutionary dynamics of lincRNA
1113 transcription in nine citrus species. Plant Journal **98,** 912-927.
1114 **Kim, D.H., and Sung, S.** (2017). Vernalization-Triggered Intragenic Chromatin Loop
1115 Formation by Long Noncoding RNAs. Dev Cell **40,** 302-312 e304.
1116 **Kindgren, P., Ard, R., Ivanov, M., and Marquardt, S.** (2018). Transcriptional read-
1117 through of the long non-coding RNA SVALKA governs plant cold acclimation. Nat
1118 Commun **9,** 4561.
1119 **Kramer, M.C., Kim, H.J., Palos, K.R., Garcia, B.A., Lyons, E., Beilstein, M.A., Nelson,**
1120 **A.D.L., and Gregory, B.D.** (2022). A Conserved Long Intergenic Non-coding
1121 RNA Containing snoRNA Sequences, lncCOBRA1, Affects Arabidopsis
1122 Germination and Development. Front Plant Sci **13,** 906603.
1123 **Li, Q.Q., Zhang, Z., Zhang, C.X., Wang, Y.L., Liu, C.B., Wu, J.C., Han, M.L., Wang,**
1124 **Q.X., and Chao, D.Y.** (2022). Phytochrome-interacting factors orchestrate
1125 hypocotyl adventitious root initiation in Arabidopsis. Development **149**.
1126 **Li, S., Yamada, M., Han, X., Ohler, U., and Benfey, P.N.** (2016). High-Resolution
1127 Expression Map of the Arabidopsis Root Reveals Alternative Splicing and
1128 lincRNA Regulation. Dev Cell **39,** 508-522.
1129 **Li, S.X., Yu, X., Lei, N., Cheng, Z.H., Zhao, P.J., He, Y.K., Wang, W.Q., and Peng, M.**
1130 (2017). Genome-wide identification and functional prediction of cold and/or
1131 drought-responsive lncRNAs in cassava. Scientific Reports **7**.
1132 **Liu, F., Marquardt, S., Lister, C., Swiezewski, S., and Dean, C.** (2010). Targeted 3'
1133 processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing.
1134 Science **327,** 94-97.
1135 **Liu, J., Wang, H., and Chua, N.H.** (2015). Long noncoding RNA transcriptome of plants.
1136 Plant Biotechnol J **13,** 319-328.
1137 **Liu, J., Jung, C., Xu, J., Wang, H., Deng, S.L., Bernad, L., Arenas-Huertero, C., and**
1138 **Chua, N.H.** (2012). Genome-Wide Analysis Uncovers Regulation of Long
1139 Intergenic Noncoding RNAs in Arabidopsis. Plant Cell **24,** 4333-4345.
1140 **Liu, Y., Tian, T., Zhang, K., You, Q., Yan, H., Zhao, N., Yi, X., Xu, W., and Su, Z.**
1141 (2018). PCSD: a plant chromatin state database. Nucleic Acids Res **46,** D1157-
1142 D1167.
1143 **Lucero, L., Ferrero, L., Fonouni-Farde, C., and Ariel, F.** (2021). Functional
1144 classification of plant long noncoding RNAs: a transcript is known by the
1145 company it keeps. New Phytologist **229,** 1251-1260.
1146 **Ma, J.C., Bai, X.T., Luo, W.C., Feng, Y.N., Shao, X.M., Bai, Q.X., Sun, S.J., Long,**
1147 **Q.M., and Wan, D.S.** (2019). Genome-Wide Identification of Long Noncoding

1148 RNAs and Their Responses to Salt Stress in Two Closely Related Poplars.
1149 Frontiers in Genetics **10**.
1150 **Mattioli, K., Volders, P.J., Gerhardinger, C., Lee, J.C., Maass, P.G., Mele, M., and**
1151 **Rinn, J.L.** (2019). High-throughput functional analysis of lncRNA core promoters
1152 elucidates rules governing tissue specificity. Genome Research **29,** 344-355.
1153 **Merelo, P., Xie, Y.K., Brand, L., Ott, F., Weigel, D., Bowman, J.L., Heisler, M.G., and**
1154 **Wenkel, S.** (2013). Genome-Wide Identification of KANADI1 Target Genes. Plos
1155 One **8**.
1156 **Mohammadin, S., Edger, P.P., Pires, J.C., and Schranz, M.E.** (2015). Positionally-
1157 conserved but sequence-diverged: identification of long non-coding RNAs in the
1158 Brassicaceae and Cleomaceae. BMC Plant Biol **15,** 217.
1159 **Moison, M., Pacheco, J.M., Lucero, L., Fonouni-Farde, C., Rodriguez-Melo, J.,**
1160 **Mansilla, N., Christ, A., Bazin, J., Benhamed, M., Ibanez, F., Crespi, M.,**
1161 **Estevez, J.M., and Ariel, F.** (2021). The lncRNA APOLO interacts with the
1162 transcription factor WRKY42 to trigger root hair cell expansion in response to
1163 cold. Mol Plant **14,** 937-948.
1164 **Moubayidin, L., Salvi, E., Giustini, L., Terpstra, I., Heidstra, R., Costantino, P., and**
1165 **Sabatini, S.** (2016). A SCARECROW-based regulatory circuit controls
1166 Arabidopsis thaliana meristem size from the root endodermis. Planta **243,** 1159-
1167 1168.
1168 **Nawrocki, E.P., and Eddy, S.R.** (2013). Infernal 1.1: 100-fold faster RNA homology
1169 searches. Bioinformatics **29,** 2933-2935.
1170 **Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., Baker,**
1171 **J.C., Grutzner, F., and Kaessmann, H.** (2014). The evolution of lncRNA
1172 repertoires and expression patterns in tetrapods. Nature **505,** 635-+.
1173 **Nelson, A.D.L., Devisetty, U.K., Palos, K., Haug-Baltzell, A.K., Lyons, E., and**
1174 **Beilstein, M.A.** (2017). Evolinc:A Tool for the Identification and Evolutionary
1175 Comparison of Long Intergenic Non-coding RNAs. Frontiers in Genetics **8**.
1176 **Nelson, A.D.L., Forsythe, E.S., Devisetty, U.K., Clausen, D.S., Haug-Batzell, A.K.,**
1177 **Meldrum, A.M.R., Frank, M.R., Lyons, E., and Beilstein, M.A.** (2016). A
1178 Genomic Analysis of Factors Driving lincRNA Diversification: Lessons from
1179 Plants. G3-Genes Genomes Genetics **6,** 2881-2891.
1180 **O'Maoileidigh, D.S., Graciet, E., and Wellmer, F.** (2014). Gene networks controlling
1181 Arabidopsis thaliana flower development. New Phytol **201,** 16-30.
1182 **Pacheco, J.M., Mansilla, N., Moison, M., Lucero, L., Gabarain, V.B., Ariel, F., and**
1183 **Estevez, J.M.** (2021). The lncRNA APOLO and the transcription factor WRKY42
1184 target common cell wall EXTENSIN encoding genes to trigger root hair cell
1185 elongation. Plant Signal Behav **16,** 1920191.
1186 **Palos, K., Dittrich, A.C.N., Yu, L.A., Brock, J.R., Railey, C.E., Wu, H.Y.L.,**
1187 **Sokolowska, E., Skirycz, A., Hsu, P.Y., Gregory, B.D., Lyons, E., Beilstein,**
1188 **M.A., and Nelson, A.D.L.** (2022). Identification and functional annotation of long
1189 intergenic non-coding RNAs in Brassicaceae. Plant Cell.
1190 **Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,**
1191 **Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos,**
1192 **A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.** (2011).
1193 Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research
1194 **12,** 2825-2830.
1195 **Pertea, G., and Pertea, M.** (2020). GFF Utilities: GffRead and GffCompare. F1000Res **9**.

1196 **Pfeiffer, A., Shi, H., Tepperman, J.M., Zhang, Y., and Quail, P.H.** (2014).
1197 Combinatorial Complexity in a Transcriptionally Centered Signaling Hub in
1198 Arabidopsis. Molecular Plant **7,** 1598-1618.
1199 **Ponting, C.P., and Haerty, W.** (2022). Genome-Wide Analysis of Human Long
1200 Noncoding RNAs: A Provocative Review. Annu Rev Genomics Hum Genet.
1201 **Qi, X., Xie, S., Liu, Y., Yi, F., and Yu, J.** (2013). Genome-wide annotation of genes and
1202 noncoding RNAs of foxtail millet in response to simulated drought stress by deep
1203 sequencing. Plant Mol Biol **83,** 459-473.
1204 **Qin, T., Zhao, H., Cui, P., Albesher, N., and Xiong, L.** (2017). A Nucleus-Localized
1205 Long Non-Coding RNA Enhances Drought and Salt Stress Tolerance. Plant
1206 Physiol **175,** 1321-1336.
1207 **Quinlan, A.R., and Hall, I.M.** (2010). BEDTools: a flexible suite of utilities for comparing
1208 genomic features. Bioinformatics **26,** 841-842.
1209 **Rai, M.I., Alam, M., Lightfoot, D.A., Gurha, P., and Afzal, A.J.** (2019). Classification
1210 and experimental identification of plant long non-coding RNAs. Genomics **111,**
1211 997-1005.
1212 **Ramskold, D., Wang, E.T., Burge, C.B., and Sandberg, R.** (2009). An abundance of
1213 ubiquitously expressed genes revealed by tissue transcriptome sequence data.
1214 PLoS Comput Biol **5,** e1000598.
1215 **Ransbotyn, V., Yeger-Lotem, E., Basha, O., Acuna, T., Verduyn, C., Gordon, M.,**
1216 **Chalifa-Caspi, V., Hannah, M.A., and Barak, S.** (2015). A combination of gene
1217 expression ranking and co-expression network analysis increases discovery rate
1218 in large-scale mutant screens for novel Arabidopsis thaliana abiotic stress genes.
1219 Plant Biotechnol J **13,** 501-513.
1220 **Ransohoff, J.D., Wei, Y., and Khavari, P.A.** (2018). The functions and unique features
1221 of long intergenic non-coding RNA. Nat Rev Mol Cell Biol **19,** 143-157.
1222 **Rigo, R., Bazin, J., Romero-Barrios, N., Moison, M., Lucero, L., Christ, A.,**
1223 **Benhamed, M., Blein, T., Huguet, S., Charon, C., Crespi, M., and Ariel, F.**
1224 (2020). The Arabidopsis lncRNA ASCO modulates the transcriptome through
1225 interaction with splicing factors. EMBO Rep **21,** e48977.
1226 **Roule, T., Crespi, M., and Blein, T.** (2022a). Regulatory long non-coding RNAs in root
1227 growth and development. Biochem Soc Trans **50,** 403-412.
1228 **Roule, T., Christ, A., Hussain, N., Huang, Y., Hartmann, C., Benhamed, M.,**
1229 **Gutierrez-Marcos, J., Ariel, F., Crespi, M., and Blein, T.** (2022b). The lncRNA
1230 MARS modulates the epigenetic reprogramming of the marneral cluster in
1231 response to ABA. Mol Plant **15,** 840-856.
1232 **Sanchita, Trivedi, P.K., and Asif, M.H.** (2020). Updates on plant long non-coding RNAs
1233 (lncRNAs): the regulatory components. Plant Cell Tissue and Organ Culture **140,**
1234 259-269.
1235 **Sarropoulos, I., Marin, R., Cardoso-Moreira, M., and Kaessmann, H.** (2019).
1236 Developmental dynamics of lncRNAs across mammalian organs and species.
1237 Nature **571,** 510-+.
1238 **Schmitz, R.J., Grotewold, E., and Stam, M.** (2022). Cis-regulatory sequences in plants:
1239 Their importance, discovery, and future challenges. Plant Cell **34,** 718-741.
1240 **Severing, E., Faino, L., Jamge, S., Busscher, M., Kuijer-Zhang, Y., Bellinazzo, F.,**
1241 **Busscher-Lange, J., Fernandez, V., Angenent, G.C., Immink, R.G.H., and**
1242 **Pajoro, A.** (2018). Arabidopsis thaliana ambient temperature responsive
1243 lncRNAs. BMC Plant Biol **18,** 145.
1244 **Shafiq, S., Li, J.R., and Sun, Q.W.** (2016). Functions of plants long non-coding RNAs.
1245 Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms **1859,** 155-162.

1246 **Shea, D.J., Nishida, N., Takada, S., Itabashi, E., Takahashi, S., Akter, A., Miyaji, N.,**
1247 **Osabe, K., Mehraj, H., Shimizu, M., Seki, M., Kakizaki, T., Okazaki, K.,**
1248 **Dennis, E.S., and Fujimoto, R.** (2019). Long noncoding RNAs in Brassica rapa
1249 L. following vernalization. Sci Rep **9,** 9302.
1250 **Shuai, P., Liang, D., Tang, S., Zhang, Z.J., Ye, C.Y., Su, Y.Y., Xia, X.L., and Yin, W.L.**
1251 (2014). Genome-wide identification and functional prediction of novel and
1252 drought-responsive lincRNAs in Populus trichocarpa. J Exp Bot **65,** 4975-4983.
1253 **Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K.,**
1254 **Clawson, H., Spieth, J., Hillier, L.W., Richards, S., Weinstock, G.M., Wilson,**
1255 **R.K., Gibbs, R.A., Kent, W.J., Miller, W., and Haussler, D.** (2005).
1256 Evolutionarily conserved elements in vertebrate, insect, worm, and yeast
1257 genomes. Genome Res **15,** 1034-1050.
1258 **Song, L., Huang, S.S.C., Wise, A., Castanon, R., Nery, J.R., Chen, H.M., Watanabe,**
1259 **M., Thomas, J., Bar-Joseph, Z., and Ecker, J.R.** (2016). A transcription factor
1260 hierarchy defines an environmental stress response network. Science **354**.
1261 **Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., Liu, Y., Chen, R., and Zhao, Y.**
1262 (2013). Utilizing sequence intrinsic composition to classify protein-coding and
1263 long non-coding transcripts. Nucleic Acids Research **41**.
1264 **Sun, Z., Nair, A., Chen, X., Prodduturi, N., Wang, J., and Kocher, J.P.** (2017).
1265 UClncR: Ultrafast and comprehensive long non-coding RNA detection from RNA-
1266 seq. Sci Rep **7,** 14196.
1267 **Szczesniak, M.W., Rosikiewicz, W., and Makalowska, I.** (2016). CANTATAdb: A
1268 Collection of Plant Long Non-Coding RNAs. Plant Cell Physiol **57,** e8.
1269 **Szczesniak, M.W., Kubiak, M.R., Wanowska, E., and Makalowska, I.** (2021).
1270 Comparative genomics in the search for conserved long noncoding RNAs.
1271 Essays Biochem **65,** 741-749.
1272 **Togninalli, M., Seren, U., Freudenthal, J.A., Monroe, J.G., Meng, D., Nordborg, M.,**
1273 **Weigel, D., Borgwardt, K., Korte, A., and Grimm, D.G.** (2020). AraPheno and
1274 the AraGWAS Catalog 2020: a major database update including RNA-Seq and
1275 knockout mutation data for Arabidopsis thaliana. Nucleic Acids Res **48,** D1063-
1276 D1068.
1277 **Tominaga-Wada, R., and Wada, T.** (2016). The ectopic localization of CAPRICE LIKE
1278 MYB3 protein in Arabidopsis root epidermis. J Plant Physiol **199,** 111-115.
1279 **Tran Vdu, T., Souiai, O., Romero-Barrios, N., Crespi, M., and Gautheret, D.** (2016).
1280 Detection of generic differential RNA processing events from RNA-seq data. RNA
1281 Biol **13,** 59-67.
1282 **Ulitsky, I.** (2016). Evolution to the rescue: using comparative genomics to understand
1283 long non-coding RNAs. Nature Reviews Genetics **17,** 601-614.
1284 **Van Bel, M., Silvestri, F., Weitz, E.M., Kreft, L., Botzki, A., Coppens, F., and**
1285 **Vandepoele, K.** (2021). PLAZA 5.0: extending the scope and power of
1286 comparative and functional genomics in plants. Nucleic Acids Res.
1287 **Van de Velde, J., Van Bel, M., Vaneechoutte, D., and Vandepoele, K.** (2016). A
1288 Collection of Conserved Noncoding Sequences to Study Gene Regulation in
1289 Flowering Plants. Plant Physiol **171,** 2586-2598.
1290 **Vandepoele, K., Quimbaya, M., Casneuf, T., De Veylder, L., and Van de Peer, Y.**
1291 (2009). Unraveling Transcriptional Control in Arabidopsis Using cis-Regulatory
1292 Elements and Coexpression Networks. Plant Physiology **150,** 535-546.
1293 **Vaneechoutte, D., and Vandepoele, K.** (2019). Curse: building expression atlases and
1294 co-expression networks from public RNA-Seq data. Bioinformatics **35,** 2880-2881.

1295 **Wang, H., Niu, Q.W., Wu, H.W., Liu, J., Ye, J., Yu, N., and Chua, N.H.** (2015a).
1296 Analysis of non-coding transcriptome in rice and maize uncovers roles of
1297 conserved lncRNAs associated with agriculture traits. Plant Journal **84,** 404-416.
1298 **Wang, H., Niu, Q.W., Wu, H.W., Liu, J., Ye, J., Yu, N., and Chua, N.H.** (2015b).
1299 Analysis of non-coding transcriptome in rice and maize uncovers roles of
1300 conserved lncRNAs associated with agriculture traits. Plant J **84,** 404-416.
1301 **Wu, H., Yang, L., and Chen, L.L.** (2017). The Diversity of Long Noncoding RNAs and
1302 Their Generation. Trends in Genetics **33,** 540-552.
1303 **Xuan, H.D., Zhang, L.Z., Liu, X.S., Han, G.M., Li, J., Li, X., Liu, A.G., Liao, M.Z., and**
1304 **Zhang, S.H.** (2015). PLNlncRbase: A resource for experimentally identified
1305 lncRNAs in plants. Gene **573,** 328-332.
1306 **Yamada, M.** (2017). Functions of long intergenic non-coding (linc) RNAs in plants. J
1307 Plant Res **130,** 67-73.
1308 **Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-**
1309 **Even, A., Horn-Saban, S., Safran, M., Domany, E., Lancet, D., and Shmueli,**
1310 **O.** (2005). Genome-wide midrange transcription profiles reveal expression level
1311 relationships in human tissue specification. Bioinformatics **21,** 650-659.
1312 **Yu, C.P., Lin, J.J., and Li, W.H.** (2016). Positional distribution of transcription factor
1313 binding sites in Arabidopsis thaliana. Sci Rep **6,** 25164.
1314 **Zhang, L., Chen, F., Zhang, X., Li, Z., Zhao, Y., Lohaus, R., Chang, X., Dong, W., Ho,**
1315 **S.Y.W., Liu, X., Song, A., Chen, J., Guo, W., Wang, Z., Zhuang, Y., Wang, H.,**
1316 **Chen, X., Hu, J., Liu, Y., Qin, Y., Wang, K., Dong, S., Liu, Y., Zhang, S., Yu, X.,**
1317 **Wu, Q., Wang, L., Yan, X., Jiao, Y., Kong, H., Zhou, X., Yu, C., Chen, Y., Li, F.,**
1318 **Wang, J., Chen, W., Chen, X., Jia, Q., Zhang, C., Jiang, Y., Zhang, W., Liu, G.,**
1319 **Fu, J., Chen, F., Ma, H., Van de Peer, Y., and Tang, H.** (2020). The water lily
1320 genome and the early evolution of flowering plants. Nature **577,** 79-84.
1321 **Zhao, Q., Li, M., Jia, Z., Liu, F., Ma, H., Huang, Y., and Song, S.** (2016). AtMYB44
1322 Positively Regulates the Enhanced Elongation of Primary Roots Induced by N-3-
1323 Oxo-Hexanoyl-Homoserine Lactone in Arabidopsis thaliana. Mol Plant Microbe
1324 Interact **29,** 774-785.
1325 **Zhao, X.Y., Li, J.R., Lian, B., Gu, H.Q., Li, Y., and Qi, Y.J.** (2018). Global identification
1326 of Arabidopsis lncRNAs reveals the regulation of MAF4 by a natural antisense
1327 RNA. Nat Commun **9**.
1328

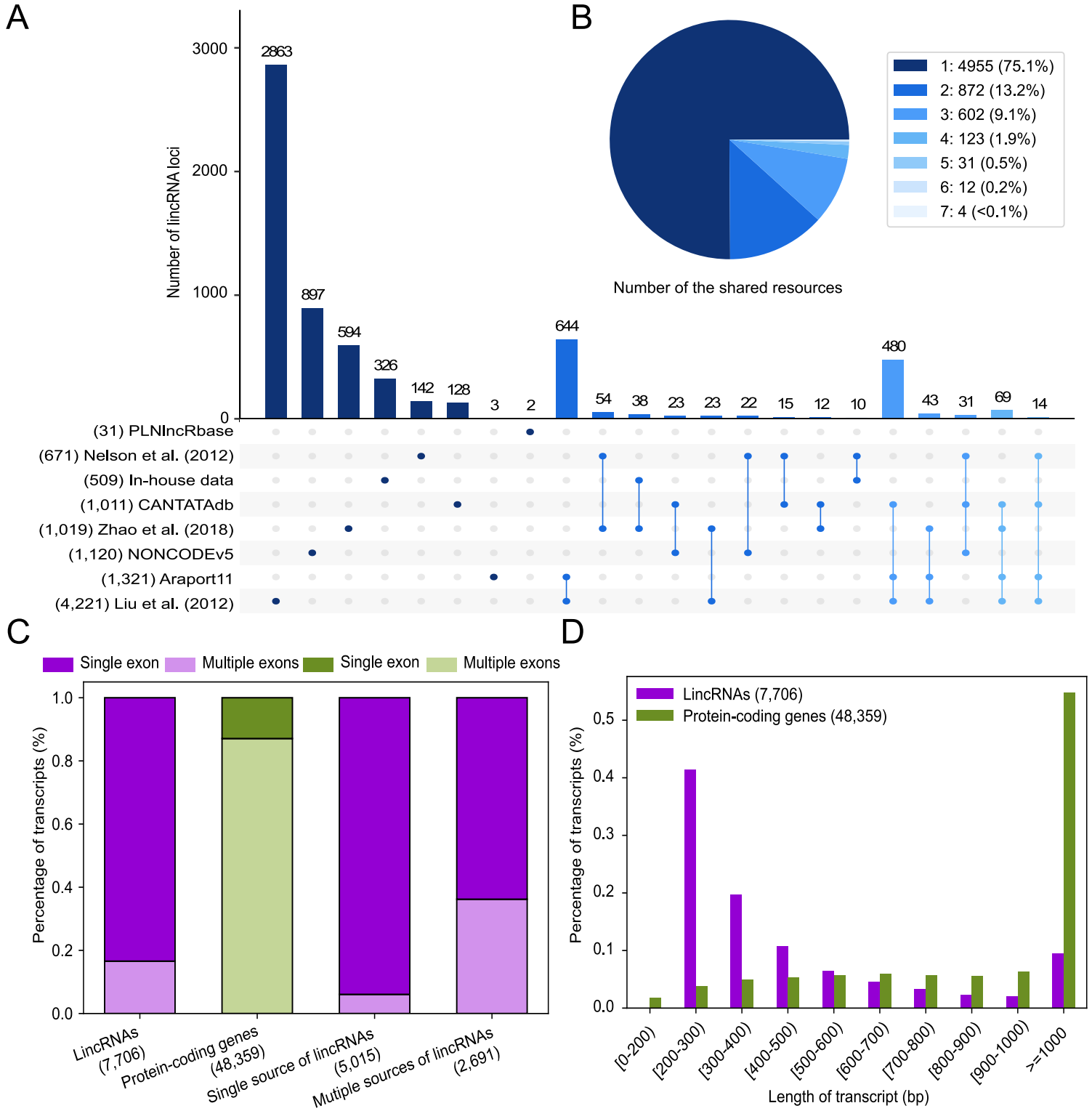



**Figure 1. Overlap and gene features of Arabidopsis lincRNA annotations. (A)** Upset plot showing the intersection of lincRNA annotation in the eight resources. Each row represents a resource, reporting in parenthesis the total number of lincRNA transcripts before merging. LincRNA annotations unique to a single resource are represented as a single circle while circles connected by lines represent the intersection of lincRNA loci shared between various resources. The bar chart indicates the number of unique lincRNA loci and intersectional lincRNA loci, displaying only intersections that contain at least ten lincRNA loci. More complex overlap patterns are not shown to maintain figure clarity. **(B)** The pie chart shows the proportion of lincRNA loci supported by one or more resources. **(C)** The distribution of exon number for all lincRNA transcripts (purple), protein-coding transcripts (green), transcripts of lincRNAs supported by single resource (purple) and multiple resources (purple). Single exon and multiple exons are shown in dark and light colors, respectively. **(D)** The distribution of transcript length for lincRNAs (purple) and protein-coding genes (green).

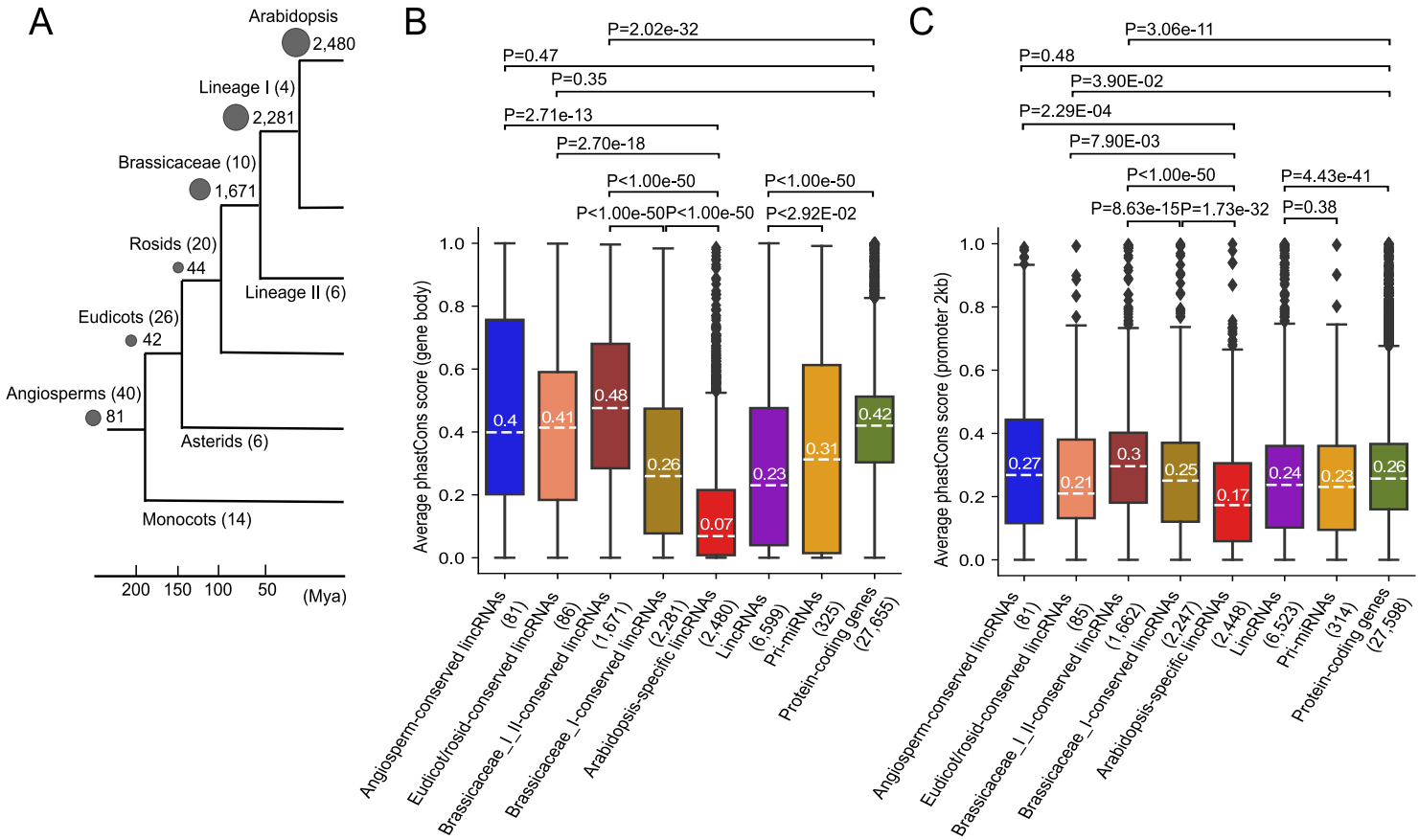



**Figure 2. Sequence conservation analysis for lincRNAs of different evolutionary age categories. (A)** Simplified species tree reporting the number of lincRNAs found for the different age categories, which are also indicated by the grey circle sizes. Numbers in parenthesis report the number of genomes included per clade to assess sequence similarity and define a lincRNA's age category. **(B)** Sequence conservation (measured by the average phastCons score over the gene body) for different lincRNA age categories and gene types (lincRNAs, pri-miRNAs, and protein-coding genes). PhastCons score ranges from 0 (not under selection) to 1 (strong negative selection). P-values for pairwise Mann–Whitney U test are shown using the horizontal lines connecting the series. **(C)** Boxplot showing the average phastCons score for promoter regions (2kb upstream of transcription start site) of different lincRNA age categories and gene types (lincRNAs, pri-miRNAs, protein-coding genes). P-values for pairwise Mann–Whitney U test are shown using the horizontal lines connecting the series.

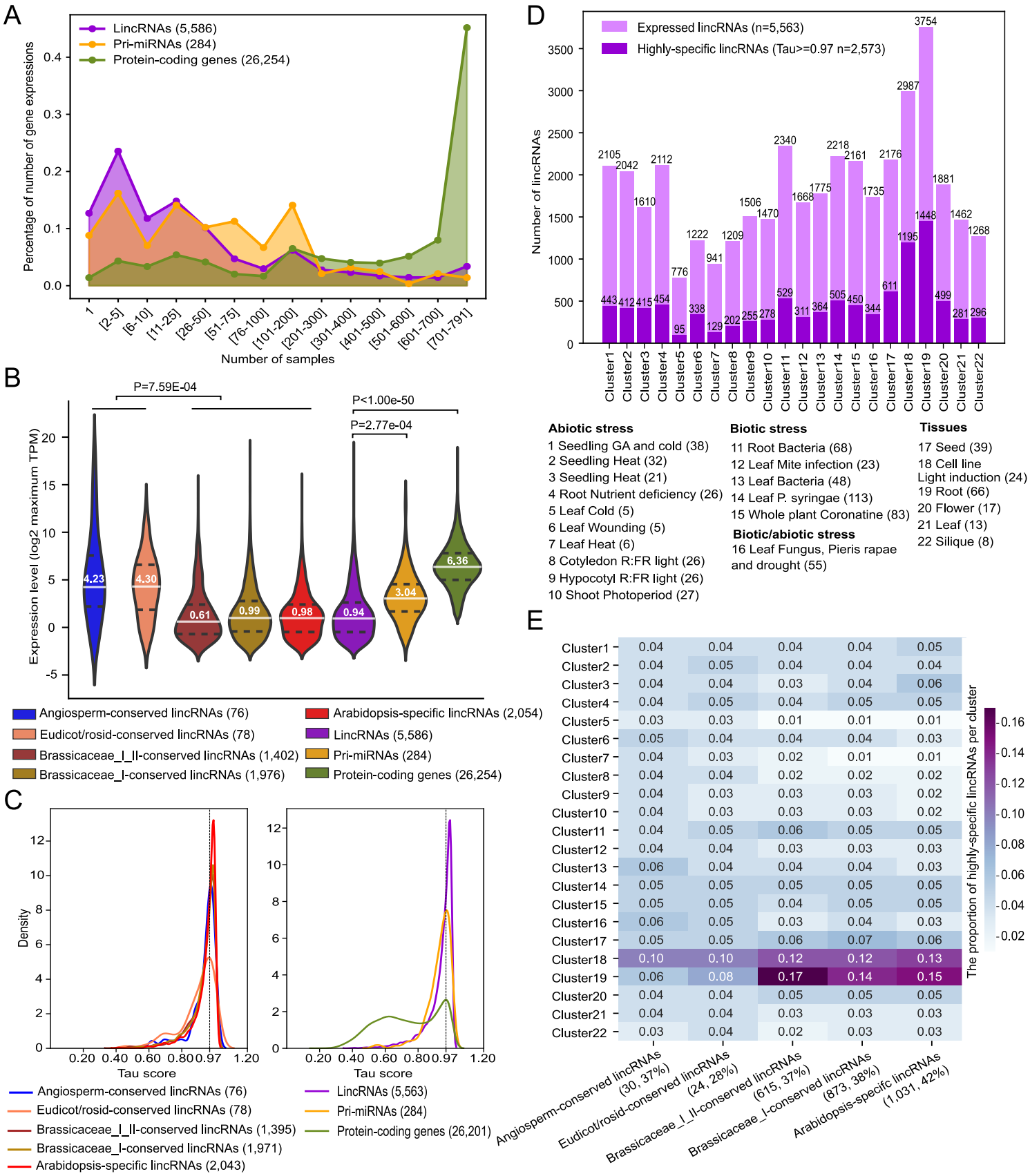



**Figure 3. Expression analysis of lincRNAs. (A)** Line chart showing the distribution of expression breadth for lincRNAs, pri-miRNAs and protein-coding genes across all samples **(B)** Distribution of the maximum TPM expression levels for different lincRNA age categories and gene types (lincRNAs, pri-miRNAs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

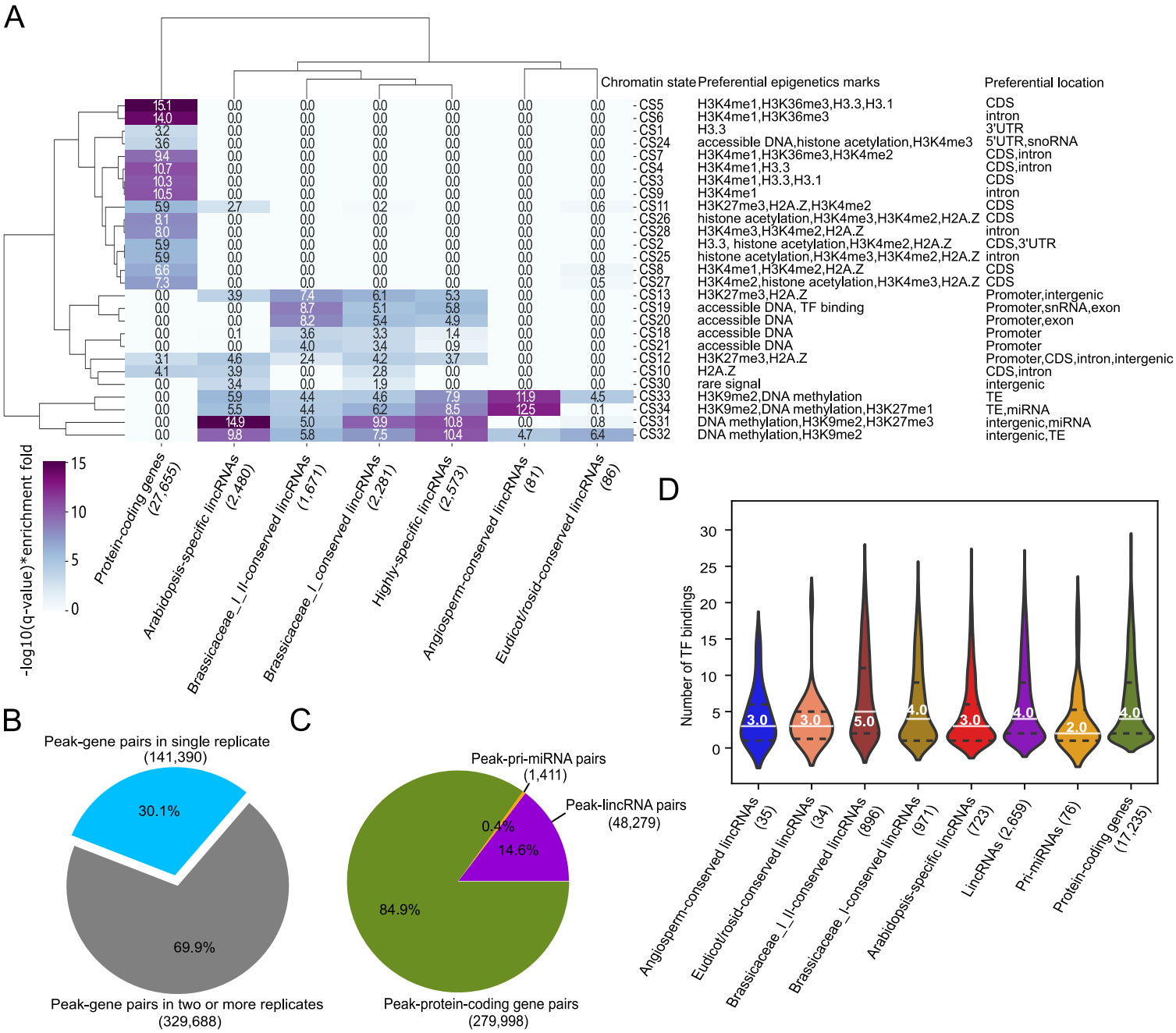



**Figure 4. Chromatin state and TF ChIP-Seq peak annotation for different gene types. (A)** Dendrogram showing the enrichment for different lincRNA gene sets (x-axis) towards different chromatin states (CS) (y-axis). The values report the product of -log10(q-value) and the enrichment fold. Only significant enrichment values are reported (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 bindings for lincRNA evolutionary age categories and gene types (lincRNAs, pri-miRNAs and protein-coding genes).

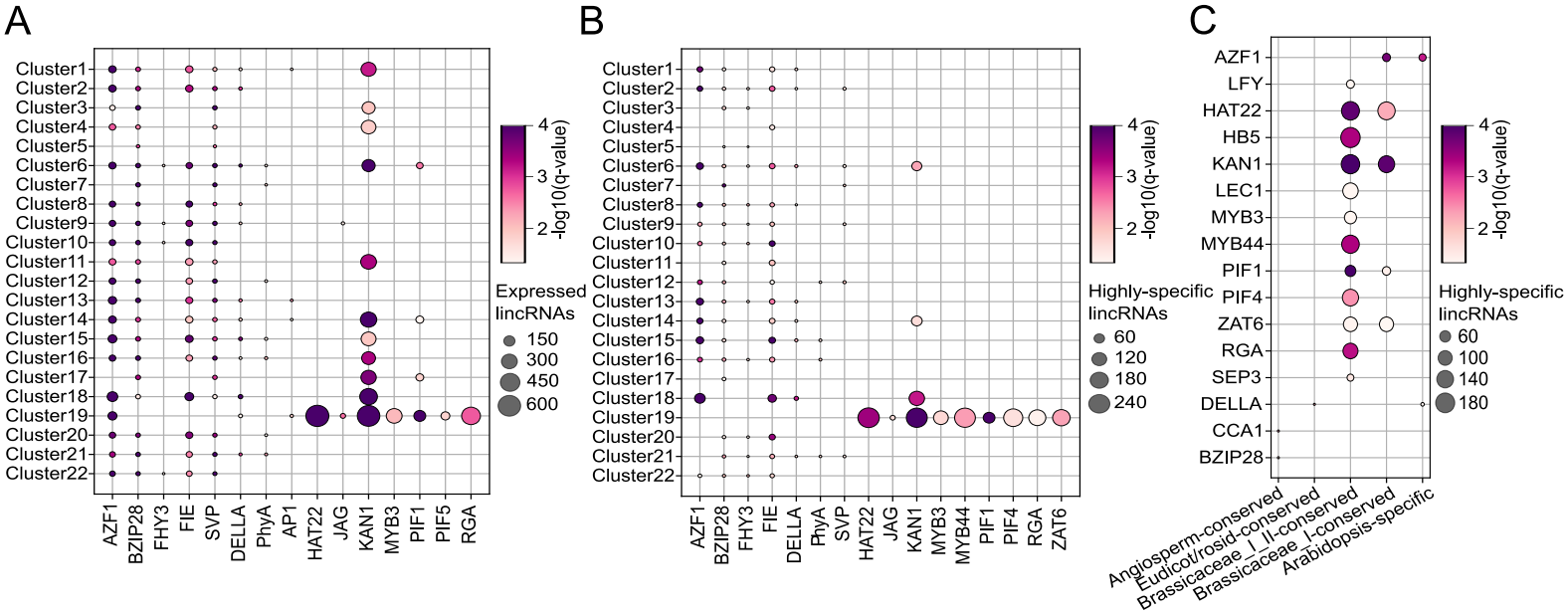



**Figure 5. Overview of TF-lincRNA regulatory interactions in different expression clusters and age categories.** The dot sizes represent the number of the lincRNAs while the color represents the statistical significance. Cluster descriptions are the same as in Figure 3D. **(A)** Bubble chart showing the enrichment of TF binding for expressed lincRNAs in different expression clusters. **(B)** Bubble chart showing the enrichment of TF binding for highly-expressed lincRNAs in different expression clusters. **(C)** Bubble chart showing the enrichment of TF binding for expressed lincRNAs in different age categories.

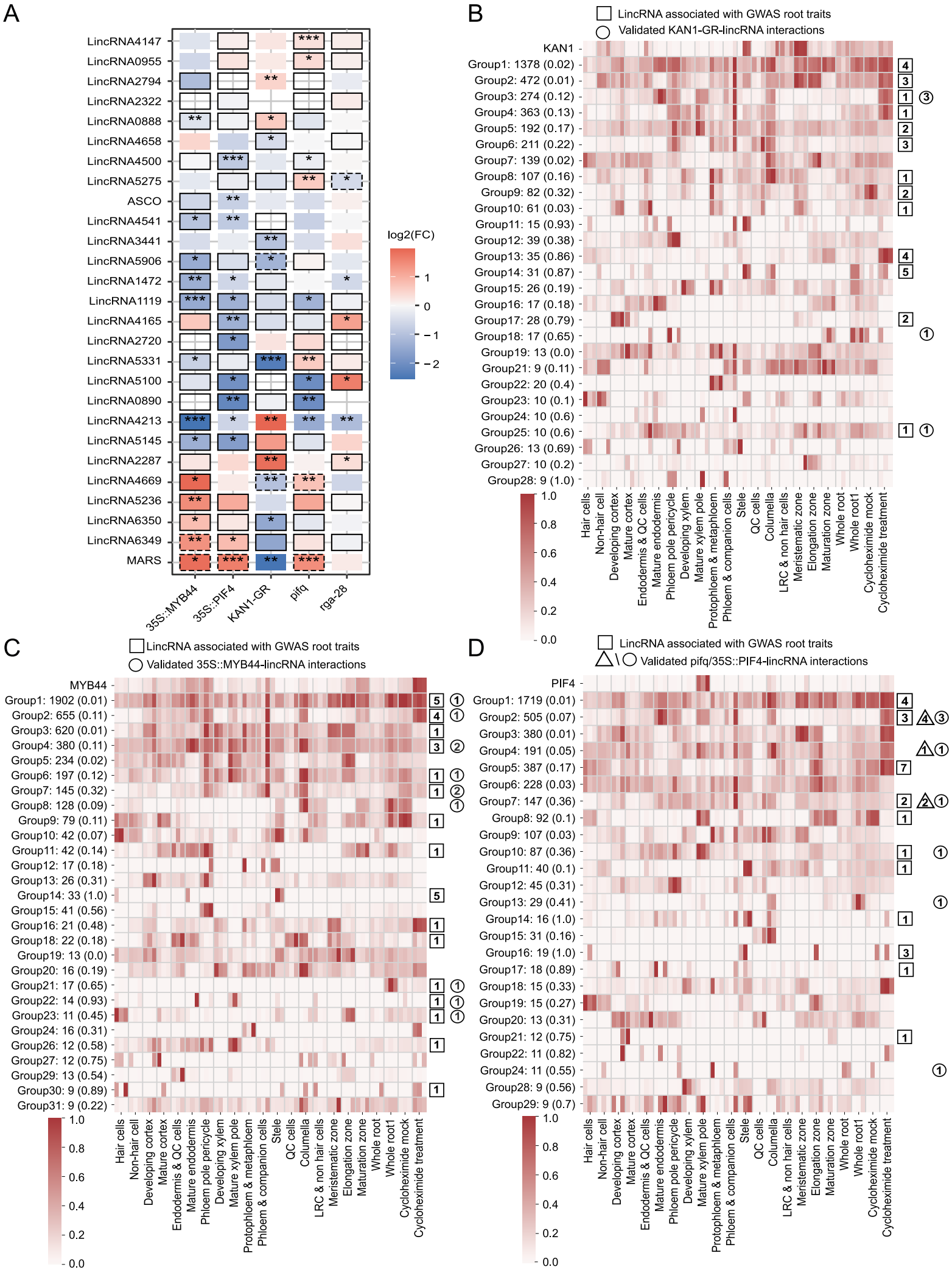

A
LincRNA4147
LincRNA0955
LincRNA2794
LincRNA2322
LincRNA0888
LincRNA4658
LincRNA4500
LincRNA5275
ASCO
LincRNA4541
LincRNA3441
LincRNA5906
LincRNA1472
LincRNA1119
LincRNA4165
LincRNA2720
LincRNA5331
LincRNA5100
LincRNA0890
LincRNA4213
LincRNA5145
LincRNA2287
LincRNA4669
LincRNA5236
LincRNA6350
LincRNA6349
MARS
35S::MYB44
35S::PIF4
KAN1-GR
pifq
rga-28
log2(FC)
1
0
−1
−2
B
LincRNA associated with GWAS root traits
Validated KAN1-GR-lincRNA interactions
KAN1
Group1: 1378 (0.02)
Group2: 472 (0.01)
Group3: 274 (0.12)
Group4: 363 (0.13)
Group5: 192 (0.17)
Group6: 211 (0.22)
Group7: 139 (0.02)
Group8: 107 (0.16)
Group9: 82 (0.32)
Group10: 61 (0.03)
Group11: 15 (0.93)
Group12: 39 (0.38)
Group13: 35 (0.86)
Group14: 31 (0.87)
Group15: 26 (0.19)
Group16: 17 (0.18)
Group17: 28 (0.79)
Group18: 17 (0.65)
Group19: 13 (0.0)
Group21: 9 (0.11)
Group22: 20 (0.4)
Group23: 10 (0.1)
Group24: 10 (0.6)
Group25: 10 (0.6)
Group26: 13 (0.69)
Group27: 10 (0.2)
Group28: 9 (1.0)
Hair cells
Non-hair cell
Developing cortex
Mature cortex
Endodermis & QC cells
Mature endodermis
Phloem pole pericycle
Developing xylem
Mature xylem pole
Protophloem & metaphloem
Phloem & companion cells
Stele
QC cells
Columella
LRC & non hair cells
Meristematic zone
Elongation zone
Maturation zone
Whole root
Whole root1
Cycloheximide mock
Cycloheximide treatment
C
LincRNA associated with GWAS root traits
Validated 35S::MYB44-lincRNA interactions
MYB44
Group1: 1902 (0.01)
Group2: 655 (0.11)
Group3: 620 (0.01)
Group4: 380 (0.11)
Group5: 234 (0.02)
Group6: 197 (0.12)
Group7: 145 (0.32)
Group8: 128 (0.09)
Group9: 79 (0.11)
Group10: 42 (0.07)
Group11: 42 (0.14)
Group12: 17 (0.18)
Group13: 26 (0.31)
Group14: 33 (1.0)
Group15: 41 (0.56)
Group16: 21 (0.48)
Group18: 22 (0.18)
Group19: 13 (0.0)
Group20: 16 (0.19)
Group21: 17 (0.65)
Group22: 14 (0.93)
Group23: 11 (0.45)
Group24: 16 (0.31)
Group26: 12 (0.58)
Group27: 12 (0.75)
Group29: 13 (0.54)
Group30: 9 (0.89)
Group31: 9 (0.22)
D
LincRNA associated with GWAS root traits
Validated pifq/35S::PIF4-lincRNA interactions
PIF4
Group1: 1719 (0.01)
Group2: 505 (0.07)
Group3: 380 (0.01)
Group4: 191 (0.05)
Group5: 387 (0.17)
Group6: 228 (0.03)
Group7: 147 (0.36)
Group8: 92 (0.1)
Group9: 107 (0.03)
Group10: 87 (0.36)
Group11: 40 (0.1)
Group12: 45 (0.31)
Group13: 29 (0.41)
Group14: 16 (1.0)
Group15: 31 (0.16)
Group16: 19 (1.0)
Group17: 18 (0.89)
Group18: 15 (0.33)
Group19: 15 (0.27)
Group20: 13 (0.31)
Group21: 12 (0.75)
Group22: 11 (0.82)
Group24: 11 (0.55)
Group28: 9 (0.56)
Group29: 9 (0.7)

**Figure 6. Experimental validation and characterization of TF-lincRNA regulatory interactions. (A)** Heat map 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* (*pif1/pif2/pif3/pif4* quadruple mutants)) vs. control wild-type lines in 14-day old roots. Expression values were determined by RT-qPCR. Asterisks indicate statistically significant differences (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$) in an unpaired two-tailed Student's t-Test (n =3). Solid boxes indicate TF ChIP peaks <2kb from the lincRNA gene. Dashed boxes indicate TF ChIP peaks were identified only in one ChIP-Seq replicate or the lincRNA is not the closest gene to the TF peak. **(B)** Heatmap showing KAN1 ChIP-Seq peak-based target genes in the cluster 19 root expression atlas. The first line shows the expression profile of the TF under investigation while the other lines show groups of target genes with similar expression. Next to the group number, the number of target genes is reported and in parenthesis the fraction of lincRNAs per group is indicated. The numbers in squares and circles report the number of GWAS root-related and RT-qPCR experimentally validated lincRNA genes, respectively. **(C)** Heatmap showing MYB44 ChIP-Seq peak-based target genes in the cluster 19 root expression atlas. Group information is the same as in panel B. **(D)** Heatmap showing PIF4 ChIP-Seq peak-based target genes in the cluster 19 root expression atlas. Group information is the same as in panel B, but here triangles and circles report the RT-qPCR experimentally validated lincRNA genes for the pifq and 35S::*PIF4* lines, respectively.

## Parsed Citations


**Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol 215, 403-410.**
Google Scholar: Author Only Title Only Author and Title

**Ariel, F., Lucero, L., Christ, A., Mammarella, M.F., Jegu, T., Veluchamy, A., Mariappan, K., Latrasse, D., Blein, T., Liu, C., Benhamed, M., and Crespi, M. (2020). R-Loop Mediated trans Action of the APOLO Long Noncoding RNA. Mol Cell 77, 1055-1065 e1054.**
Google Scholar: Author Only Title Only Author and Title

**Bardou, F., Ariel, F., Simpson, C.G., Romero-Barrios, N., Laporte, P., Balzergue, S., Brown, J.W., and Crespi, M. (2014). Long noncoding RNA modulates alternative splicing regulators in Arabidopsis. Dev Cell 30, 166-176.**
Google Scholar: Author Only Title Only Author and Title

**Bazin, J., Romero, N., Rigo, R., Charon, C., Blein, T., Ariel, F., and Crespi, M. (2018). Nuclear Speckle RNA Binding Proteins Remodel Alternative Splicing and the Non-coding Arabidopsis Transcriptome to Regulate a Cross-Talk Between Auxin and Immune Responses. Frontiers in Plant Science 9.**
Google Scholar: Author Only Title Only Author and Title

**Beilstein, M.A., Nagalingum, N.S., Clements, M.D., Manchester, S.R., and Mathews, S. (2010). Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana. Proc Natl Acad Sci U S A 107, 18724-18728.**
Google Scholar: Author Only Title Only Author and Title

**Bhogireddy, S., Mangrauthia, S.K., Kumar, R., Pandey, A.K., Singh, S., Jain, A., Budak, H., Varshney, R.K., and Kudapa, H. (2021). Regulatory non-coding RNAs: a new frontier in regulation of plant biology. Functional & Integrative Genomics 21, 313-330.**
Google Scholar: Author Only Title Only Author and Title

**Blein, T., Balzergue, C., Roule, T., Gabriel, M., Scalisi, L., Francois, T., Sorin, C., Christ, A., Godon, C., Delannoy, E., Martin-Magniette, M.L., Nussaume, L., Hartmann, C., Gautheret, D., Desnos, T., and Crespi, M. (2020). Landscape of the Noncoding Transcriptome Response of Two Arabidopsis Ecotypes to Phosphate Starvation. Plant Physiol 183, 1058-1072.**
Google Scholar: Author Only Title Only Author and Title

**Bu, D.C., Luo, H.T., Jiao, F., Fang, S.S., Tan, C.F., Liu, Z.Y., and Zhao, Y. (2015). Evolutionary annotation of conserved long non-coding RNAs in major mammalian species. Science China-Life Sciences 58, 787-798.**
Google Scholar: Author Only Title Only Author and Title

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

**Chen, L., Zhu, Q.H., and Kaufmann, K. (2020). Long non-coding RNAs in plants: emerging modulators of gene activity in development and stress responses. Planta 252, 92.**
Google Scholar: Author Only Title Only Author and Title

**Chen, Q.S., Liu, K., Yu, R., Zhou, B.L., Huang, P.P., Cao, Z.X., Zhou, Y.Q., and Wang, J.H. (2021). From "Dark Matter" to "Star": Insight Into the Regulation Mechanisms of Plant Functional Long Non-Coding RNAs. Frontiers in Plant Science 12.**
Google Scholar: Author Only Title Only Author and Title

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

**Chow, C.N., Lee, T.Y., Hung, Y.C., Li, G.Z., Tseng, K.C., Liu, Y.H., Kuo, P.L., Zheng, H.Q., and Chang, W.C. (2019). PlantPAN3.0: a new and updated resource for reconstructing transcriptional regulatory networks from ChIP-seq experiments in plants. Nucleic Acids Research 47, D1155-D1163.**
Google Scholar: Author Only Title Only Author and Title

**Deng, F., Zhang, X., Wang, W., Yuan, R., and Shen, F. (2018). Identification of Gossypium hirsutum long non-coding RNAs (lncRNAs) under salt stress. BMC Plant Biol 18, 23.**
Google Scholar: Author Only Title Only Author and Title

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

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

**Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas,**

**A., Salazar, G.A., Tate, J., and Bateman, A. (2016). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44, D279-285.**
Google Scholar: Author Only Title Only Author and Title

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

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

**Hawker, N.P., and Bowman, J.L. (2004). Roles for Class III HD-Zip and KANADI genes in Arabidopsis root development. Plant Physiol 135, 2261-2270.**
Google Scholar: Author Only Title Only Author and Title

**Hazarika, R.R., Serra, M., Zhang, Z., Zhang, Y., Schmitz, R.J., and Johannes, F. (2022). Molecular properties of epimutation hotspots. Nat Plants 8, 146-156.**
Google Scholar: Author Only Title Only Author and Title

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

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

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

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

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

**Hupalo, D., and Kern, A.D. (2013). Conservation and functional element discovery in 20 angiosperm plant genomes. Mol Biol Evol 30, 1729-1744.**
Google Scholar: Author Only Title Only Author and Title

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

**Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. (2006). MicroRNAs and their regulatory roles in plants. Annual Review of Plant Biology 57, 19-53.**
Google Scholar: Author Only Title Only Author and Title

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

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

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

**Kim, D.H., and Sung, S. (2017). Vernalization-Triggered Intragenic Chromatin Loop Formation by Long Noncoding RNAs. Dev Cell 40, 302-312 e304.**
Google Scholar: Author Only Title Only Author and Title

**Kindgren, P., Ard, R., Ivanov, M., and Marquardt, S. (2018). Transcriptional read-through of the long non-coding RNA SVALKA governs plant cold acclimation. Nat Commun 9, 4561.**
Google Scholar: Author Only Title Only Author and Title

**Kramer, M.C., Kim, H.J., Palos, K.R., Garcia, B.A., Lyons, E., Beilstein, M.A., Nelson, A.D.L., and Gregory, B.D. (2022). A Conserved Long Intergenic Non-coding RNA Containing snoRNA Sequences, lncCOBRA1, Affects Arabidopsis Germination and Development. Front Plant Sci 13, 906603.**
Google Scholar: Author Only Title Only Author and Title

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

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

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

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

**Liu, J., Wang, H., and Chua, N.H. (2015). Long noncoding RNA transcriptome of plants. Plant Biotechnol J 13, 319-328.**
Google Scholar: Author Only Title Only Author and Title

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

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

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

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

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

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

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

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

**Moubayidin, L., Salvi, E., Giustini, L., Terpstra, I., Heidstra, R., Costantino, P., and Sabatini, S. (2016). A SCARECROW-based regulatory circuit controls Arabidopsis thaliana meristem size from the root endodermis. Planta 243, 1159-1168.**
Google Scholar: Author Only Title Only Author and Title

**Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29, 2933-2935.**
Google Scholar: Author Only Title Only Author and Title

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

**Nelson, A.D.L., Devisetty, U.K., Palos, K., Haug-Baltzell, A.K., Lyons, E., and Beilstein, M.A. (2017). Evolinc:A Tool for the Identification and Evolutionary Comparison of Long Intergenic Non-coding RNAs. Frontiers in Genetics 8.**
Google Scholar: Author Only Title Only Author and Title

**Nelson, A.D.L., Forsythe, E.S., Devisetty, U.K., Clausen, D.S., Haug-Batzell, A.K., Meldrum, A.M.R., Frank, M.R., Lyons, E., and Beilstein, M.A. (2016). A Genomic Analysis of Factors Driving lincRNA Diversification: Lessons from Plants. G3-Genes Genomes Genetics 6, 2881-2891.**
Google Scholar: Author Only Title Only Author and Title

**O'Maoileidigh, D.S., Graciet, E., and Wellmer, F. (2014). Gene networks controlling Arabidopsis thaliana flower development. New Phytol 201, 16-30.**
Google Scholar: Author Only Title Only Author and Title

**Pacheco, J.M., Mansilla, N., Moison, M., Lucero, L., Gabarain, V.B., Ariel, F., and Estevez, J.M. (2021). The lncRNA APOLO and the transcription factor WRKY42 target common cell wall EXTENSIN encoding genes to trigger root hair cell elongation. Plant Signal Behav 16, 1920191.**
Google Scholar: Author Only Title Only Author and Title

**Palos, K., Dittrich, A.C.N., Yu, L.A., Brock, J.R., Railey, C.E., Wu, H.Y.L., Sokolowska, E., Skirycz, A., Hsu, P.Y., Gregory, B.D., Lyons, E., Beilstein, M.A., and Nelson, A.D.L. (2022). Identification and functional annotation of long intergenic non-coding RNAs in Brassicaceae. Plant Cell.**
Google Scholar: Author Only Title Only Author and Title

**Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825-2830.**
Google Scholar: Author Only Title Only Author and Title

**Pertea, G., and Pertea, M. (2020). GFF Utilities: GffRead and GffCompare. F1000Res 9.**
Google Scholar: Author Only Title Only Author and Title

**Pfeiffer, A., Shi, H., Tepperman, J.M., Zhang, Y., and Quail, P.H. (2014). Combinatorial Complexity in a Transcriptionally Centered Signaling Hub in Arabidopsis. Molecular Plant 7, 1598-1618.**
Google Scholar: Author Only Title Only Author and Title

**Ponting, C.P., and Haerty, W. (2022). Genome-Wide Analysis of Human Long Noncoding RNAs: A Provocative Review. Annu Rev Genomics Hum Genet.**
Google Scholar: Author Only Title Only Author and Title

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

**Qin, T., Zhao, H., Cui, P., Albesher, N., and Xiong, L. (2017). A Nucleus-Localized Long Non-Coding RNA Enhances Drought and Salt Stress Tolerance. Plant Physiol 175, 1321-1336.**
Google Scholar: Author Only Title Only Author and Title

**Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842.**
Google Scholar: Author Only Title Only Author and Title

**Rai, M.I., Alam, M., Lightfoot, D.A., Gurha, P., and Afzal, A.J. (2019). Classification and experimental identification of plant long non-coding RNAs. Genomics 111, 997-1005.**
Google Scholar: Author Only Title Only Author and Title

**Ramskold, D., Wang, E.T., Burge, C.B., and Sandberg, R. (2009). An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput Biol 5, e1000598.**
Google Scholar: Author Only Title Only Author and Title

**Ransbotyn, V., Yeger-Lotem, E., Basha, O., Acuna, T., Verduyn, C., Gordon, M., Chalifa-Caspi, V., Hannah, M.A., and Barak, S. (2015). A combination of gene expression ranking and co-expression network analysis increases discovery rate in large-scale mutant screens for novel Arabidopsis thaliana abiotic stress genes. Plant Biotechnol J 13, 501-513.**
Google Scholar: Author Only Title Only Author and Title

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

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

**Roule, T., Crespi, M., and Blein, T. (2022a). Regulatory long non-coding RNAs in root growth and development. Biochem Soc Trans 50, 403-412.**
Google Scholar: Author Only Title Only Author and Title

**Roule, T., Christ, A., Hussain, N., Huang, Y., Hartmann, C., Benhamed, M., Gutierrez-Marcos, J., Ariel, F., Crespi, M., and Blein, T. (2022b). The lncRNA MARS modulates the epigenetic reprogramming of the marneral cluster in response to ABA. Mol Plant 15, 840-856.**
Google Scholar: Author Only Title Only Author and Title

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

**Sarropoulos, I., Marin, R., Cardoso-Moreira, M., and Kaessmann, H. (2019). Developmental dynamics of lncRNAs across mammalian organs and species. Nature 571, 510-+.**
Google Scholar: Author Only Title Only Author and Title

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

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

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

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

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

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

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

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

**Sun, Z., Nair, A., Chen, X., Prodduturi, N., Wang, J., and Kocher, J.P. (2017). UCIncR: Ultrafast and comprehensive long non-coding RNA detection from RNA-seq. Sci Rep 7, 14196.**
Google Scholar: Author Only Title Only Author and Title

**Szczesniak, M.W., Rosikiewicz, W., and Makalowska, I. (2016). CANTATAdb: A Collection of Plant Long Non-Coding RNAs. Plant Cell Physiol 57, e8.**
Google Scholar: Author Only Title Only Author and Title

**Szczesniak, M.W., Kubiak, M.R., Wanowska, E., and Makalowska, I. (2021). Comparative genomics in the search for conserved**

**long noncoding RNAs. Essays Biochem 65, 741-749.**
Google Scholar: Author Only Title Only Author and Title

**Togninalli, M., Seren, U., Freudenthal, J.A., Monroe, J.G., Meng, D., Nordborg, M., Weigel, D., Borgwardt, K., Korte, A., and Grimm, D.G. (2020). AraPheno and the AraGWAS Catalog 2020: a major database update including RNA-Seq and knockout mutation data for Arabidopsis thaliana. Nucleic Acids Res 48, D1063-D1068.**
Google Scholar: Author Only Title Only Author and Title

**Tominaga-Wada, R., and Wada, T. (2016). The ectopic localization of CAPRICE LIKE MYB3 protein in Arabidopsis root epidermis. J Plant Physiol 199, 111-115.**
Google Scholar: Author Only Title Only Author and Title

**Tran Vdu, T., Souiai, O., Romero-Barrios, N., Crespi, M., and Gautheret, D. (2016). Detection of generic differential RNA processing events from RNA-seq data. RNA Biol 13, 59-67.**
Google Scholar: Author Only Title Only Author and Title

**Ulitsky, I. (2016). Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. Nature Reviews Genetics 17, 601-614.**
Google Scholar: Author Only Title Only Author and Title

**Van Bel, M., Silvestri, F., Weitz, E.M., Kreft, L., Botzki, A., Coppens, F., and Vandepoele, K. (2021). PLAZA 5.0: extending the scope and power of comparative and functional genomics in plants. Nucleic Acids Res.**
Google Scholar: Author Only Title Only Author and Title

**Van de Velde, J., Van Bel, M., Vaneechoutte, D., and Vandepoele, K. (2016). A Collection of Conserved Noncoding Sequences to Study Gene Regulation in Flowering Plants. Plant Physiol 171, 2586-2598.**
Google Scholar: Author Only Title Only Author and Title

**Vandepoele, K., Quimbaya, M., Casneuf, T., De Veylder, L., and Van de Peer, Y. (2009). Unraveling Transcriptional Control in Arabidopsis Using cis-Regulatory Elements and Coexpression Networks. Plant Physiology 150, 535-546.**
Google Scholar: Author Only Title Only Author and Title

**Vaneechoutte, D., and Vandepoele, K. (2019). Curse: building expression atlases and co-expression networks from public RNA-Seq data. Bioinformatics 35, 2880-2881.**
Google Scholar: Author Only Title Only Author and Title

**Wang, H., Niu, Q.W., Wu, H.W., Liu, J., Ye, J., Yu, N., and Chua, N.H. (2015a). Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved lncRNAs associated with agriculture traits. Plant Journal 84, 404-416.**
Google Scholar: Author Only Title Only Author and Title

**Wang, H., Niu, Q.W., Wu, H.W., Liu, J., Ye, J., Yu, N., and Chua, N.H. (2015b). Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved lncRNAs associated with agriculture traits. Plant J 84, 404-416.**
Google Scholar: Author Only Title Only Author and Title

**Wu, H., Yang, L., and Chen, L.L. (2017). The Diversity of Long Noncoding RNAs and Their Generation. Trends in Genetics 33, 540-552.**
Google Scholar: Author Only Title Only Author and Title

**Xuan, H.D., Zhang, L.Z., Liu, X.S., Han, G.M., Li, J., Li, X., Liu, A.G., Liao, M.Z., and Zhang, S.H. (2015). PLNlncRbase: A resource for experimentally identified lncRNAs in plants. Gene 573, 328-332.**
Google Scholar: Author Only Title Only Author and Title

**Yamada, M. (2017). Functions of long intergenic non-coding (linc) RNAs in plants. J Plant Res 130, 67-73.**
Google Scholar: Author Only Title Only Author and Title

**Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E., Lancet, D., and Shmueli, O. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics 21, 650-659.**
Google Scholar: Author Only Title Only Author and Title

**Yu, C.P., Lin, J.J., and Li, W.H. (2016). Positional distribution of transcription factor binding sites in Arabidopsis thaliana. Sci Rep 6, 25164.**
Google Scholar: Author Only Title Only Author and Title

**Zhang, L., Chen, F., Zhang, X., Li, Z., Zhao, Y., Lohaus, R., Chang, X., Dong, W., Ho, S.Y.W., Liu, X., Song, A., Chen, J., Guo, W., Wang, Z., Zhuang, Y., Wang, H., Chen, X., Hu, J., Liu, Y., Qin, Y., Wang, K., Dong, S., Liu, Y., Zhang, S., Yu, X., Wu, Q., Wang, L., Yan, X., Jiao, Y., Kong, H., Zhou, X., Yu, C., Chen, Y., Li, F., Wang, J., Chen, W., Chen, X., Jia, Q., Zhang, C., Jiang, Y., Zhang, W., Liu, G., Fu, J., Chen, F., Ma, H., Van de Peer, Y., and Tang, H. (2020). The water lily genome and the early evolution of flowering plants. Nature 577, 79-84.**
Google Scholar: Author Only Title Only Author and Title

**Zhao, Q., Li, M., Jia, Z., Liu, F., Ma, H., Huang, Y., and Song, S. (2016). AtMYB44 Positively Regulates the Enhanced Elongation of Primary Roots Induced by N-3-Oxo-Hexanoyl-Homoserine Lactone in Arabidopsis thaliana. Mol Plant Microbe Interact 29, 774-785.**
Google Scholar: Author Only Title Only Author and Title

**Zhao, X.Y., Li, J.R., Lian, B., Gu, H.Q., Li, Y., and Qi, Y.J. (2018). Global identification of Arabidopsis lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. Nat Commun 9.**
Google Scholar: Author Only Title Only Author and Title