Loss of Polycomb proteins CLF and LHP1 leads to excessive RNA degradation in Arabidopsis

David Séré, Océane Cassan, Fanny Bellegarde, Cécile Fizames, Jossia Boucherez, Geoffrey Schivre, Jacinthe Azevedo, Thierry Lagrange, Alain Gojon, Antoine Martin

Loss of Polycomb proteins CLF and LHP1 leads to excessive RNA degradation in Arabidopsis

David Séré[1], Océane Cassan[1], Fanny Bellegarde[1,3], Cécile Fizames[1], Jossia Boucherez[1], Geoffrey Schivre[1], Jacinthe Azevedo[2], Thierry Lagrange[2], Alain Gojon[1], Antoine Martin[1*]

[1] IPSiM, Univ Montpellier, CNRS, INRAE, Institut Agro, Montpellier, France

[2] CNRS, LGDP UMR5096, Université de Perpignan, 66860 Perpignan, France

[3] Graduate School of Bioagricultural Sciences, Nagoya University, Aichi 464-8601, Japan.

[*]corresponding author: Antoine Martin, antoine.martin@cnrs.fr

Highlight

The function of Polycomb-group proteins is essential to regulate the homeostasis of RNA degradation, and ensure proper gene expression levels.

Abstract

Polycomb-group (PcG) proteins are major chromatin complexes that regulate gene expression, mainly described as repressors keeping genes in a transcriptional silent state during development. Recent studies have nonetheless suggested that PcG proteins might have additional functions, including on targeting active genes or independently of gene expression regulation. However, the reasons for the implication of PcG and their associated chromatin marks on active genes are still largely unknown. Here, we report that combining mutations for CURLY LEAF (CLF) and LIKE HETEROCHROMATIN PROTEIN1 (LHP1), two Arabidopsis PcG proteins, results in the deregulation of the expression of active genes that are targeted by PcG proteins or enriched in associated chromatin marks. We show that this deregulation is associated with accumulation of small RNAs corresponding to massive degradation of active gene transcripts. We demonstrate that transcriptionally active genes and especially those targeted by PcG are prone to RNA degradation, even though deregulation of RNA degradation following the loss of function of PcG proteins is likely not mediated by a PcG-mediated chromatin environment. Therefore, we highlight that PcG function is essential to maintain an accurate level of RNA degradation to ensure a correct expression level of genes.

Keywords

Arabidopsis, chromatin, nutrition, gene expression, RNA degradation, Polycomb, roots.

Introduction

Accurate expression of genes is the result of a range of transcriptional and post-transcriptional mechanisms. At the chromatin level, a large number of regulatory pathways and their associated factors have been identified, and are major determinants of the regulation of genome expression. Among them, the Polycomb-Group (PcG) proteins have been largely described for their role in the establishment of genomic programs in eukaryotes (Mozgova and Hennig, 2015; Schuettengruber *et al.*, 2017). In plants, PcG regulation is ensured by two different complexes (Polycomb Repressive Complexes), PRC1 and PCR2 (Mozgova and Hennig, 2015). The enzymatic function of PRC1 is to catalyze the monoubiquitination of the lysine 119 on histone H2A (H2AK119ub), while PRC2 catalyzes the trimethylation of the lysine 27 on histone H3 (H3K27me3) (Forderer *et al.*, 2016). The function of these two chromatin modifications is to reduce the expression of genes and to maintain their expression at low levels. Therefore, H2AK119ub, H3K27me3, PRC1 and PRC2 are preferentially associated with genes showing low expression (Roudier *et al.*, 2011; Sequeira-Mendes *et al.*, 2014; Turck *et al.*, 2007; Zhou *et al.*, 2017a). In Arabidopsis, PRC1 is composed of several typical E3 ubiquitin ligases, BMI1A/B/C and RING1A/B, associated with the plant-specific factors LHP1 and EMF1 (Wang and Shen, 2018). Arabidopsis PRC2 is organized around three H3K27 trimethyl-transferase enzymes, among them CLF, and to a lesser extent SWN, are responsible for H3K27me3 deposition in vegetative tissues like shoots and roots (Mozgova *et al.*, 2015). A large number of studies, in animals and in plants, have drawn a schematic cooperative model for the hierarchical actions of PRC1 and PRC2. This model proposes that PRC2 might first catalyze H3K27me3 on target genes, which will be recognized by H3K27me3 readers such as LHP1 or other PRC1-associated factors (Li *et al.*, 2018; Turck *et al.*, 2007). This cooperative model has been largely described and yet several recent studies have challenged it, proposing an opposite hierarchy between PRC2 and PRC1, as well as independent functions and genomic targets (Merini and Calonje, 2015; Zhou *et al.*, 2017a). Nevertheless, several pieces of evidence strongly support the hypothesis of a functional link between PRC2 and PRC1, and especially between CLF and LHP1. First, LHP1 physically interacts with several members of PRC2

(Derkacheva *et al.*, 2013; Hecker *et al.*, 2015), suggesting that it acts as a bridge between PRC2 and PRC1. In addition, *clf* and *lhp1* mutants have similar developmental phenotypes and significantly overlapping transcriptome, supporting a common function (Veluchamy *et al.*, 2016; Wang *et al.*, 2016; Zhou *et al.*, 2017b). Taken together, these observations mean that PRC1 and PRC2 cooperate, to a certain extent, to down-regulate the expression of genes, and to keep their expression silenced. However, several studies have demonstrated or suggested more diverse functions for PcG proteins than the strict transcriptional repression of genes. First, recent reports have illustrated PcG functions that are not restricted to genes with low expression. In mammalian cells for instance, a non-canonical PRC1 targets preferentially active loci, associated with metabolic functions (van den Boom *et al.*, 2016). Likewise, it has been demonstrated in Drosophila that PRC1 could regulate the expression of a subset of active genes, with a large part of this regulation being certainly direct (Pherson *et al.*, 2017). In Arabidopsis, a subset of H2AUb-marked genes corresponds to transcriptionally active genes (Zhou *et al.*, 2017a). Interestingly these genes are also associated with metabolic functions, but the impact of PcG proteins regulation on them has not been analyzed. Another study in Arabidopsis described that *NRT2.1*, which is a highly expressed gene coding for a major and essential root nitrate transporter, is constantly marked by H3K27me3 even under transcriptionally permissive conditions (Bellegarde *et al.*, 2018). Loss of CLF function leads to a reduction of H3K27me3 levels at the *NRT2.1* locus, increasing its mRNA levels without altering its expression pattern (Bellegarde *et al.*, 2018). On another hand, non-transcriptional functions have been also described for PcG proteins. Indeed, different reports have demonstrated that PcG proteins can be directly implicated in cellular processes independently of the regulation of gene expression, including the regulation of cell cycle progression, or the regulation of alternative splicing (Brien *et al.*, 2015; Gonzalez *et al.*, 2015; Lecona *et al.*, 2013; Mohd-Sarip *et al.*, 2012). Therefore, growing evidence show that PcG proteins may have emerging functions that are not strictly associated with repression of gene expression. However, even though the presence of PcG proteins and their associated chromatin marks on transcriptionally active loci have been clearly demonstrated, the implication in their regulation is still elusive and deserves

further investigation. Here, we used *NRT2.1* as a model gene to further explore the role of PcG proteins in the regulation of transcriptionally active genes in Arabidopsis. We showed that in addition to being controlled by PRC2, *NRT2.1* is also regulated by PRC1 members, including LHP1. Concurrent mutations in CLF and LHP1 lead to an unexpected down-regulation of *NRT2.1* expression, associated with the accumulation of sRNAs resulting from mRNA degradation. We demonstrated that over-accumulation of sRNAs in a double *clf lhp1* mutant concerns a large set of transcriptionally active genes, that are consequently deregulated. We finally observe that mRNA degradation globally increases in *clf lhp1* mutant, likely corresponding to a misregulation of actors of RNA degradation machinery.

Material and Methods

Plant material and growth conditions

The *Arabidopsis thaliana* accession used in this study was *Col-0*. Mutant alleles and transgenic plants used in this study are *clf-29* (Xu and Shen, 2008), *lhp1-4* (Takada and Goto, 2003), *ProNRT2.1:LUC* (Girin *et al.*, 2010), *bmi1a*, *bmi1b*, *bmi1c* (Merini *et al.*, 2017). Experiments were performed using roots from 5 weeks-old seedlings grown hydroponically under a short-day photoperiod (8 h light and 16 h dark) and supplied with MS/2 solution containing 0,3 mM KNO_3.

Expression analysis

Root samples were frozen in liquid nitrogen and tissues were disrupted for 1 min at 30 s−1 in a Retsch mixer mill MM301 homogenizer (Retsch, Haan, Germany). Total RNA was extracted using TRI REAGENT (*MRC*). Subsequently 500 ng of total RNA were treated with DNase (DNase I; SIGMA-ALDRICH, St. Louis, MO, USA) following the manufacturer's instructions. Reverse transcription was achieved with M-MLV reverse transcriptase (RNase H minus, Point Mutant, *Promega*) using 1 µL of an anchored oligo$(dT)_{20}$ primer (50 µM) for mRNA, or specific complementary primers (50 µM) with stem loop secondary structures for miRNAs (Varkonyi-Gasic *et al.*, 2007). Accumulation of transcripts

was measured by qRT-PCR (LightCycler 480, *Roche Diagnostics*) using the SYBR^{R} Premix Ex Taq^{TM} (*TaKaRa*), according to the manufacturer's instructions with 1 µL of cDNA in a total reaction volume of 10 µL. Gene expression was normalized using *UBQ10* or *ACT2* as internal standards and using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001). Sequences of primers used in qPCR for gene expression and stem-loop analysis are listed in Supplemental dataset.

Root nitrate influx

NO_3^- influx was assayed using 0.2 mM $^{15}NO_3$. Roots were then dried at 70°C for 48 h, and samples were analyzed for total N and atom% ^{15}N using a continuous flow isotope ratio mass spectrometer coupled with a C/N elemental analyzer (model Euroflash; Eurovector, Pavia, Italy) as previously described (Jacquot *et al.*, 2020).

Transcriptome profiling

Genome-wide expression analyses were performed on using Arabidopsis Affymetrix Gene 1.1 ST array strips designed to measure whole transcript accumulation of 28,501 genes (or transcripts clusters), based on 600,941 probes designed based on TAIR10 genome annotation. Biotin labeled and fragmented cRNAs were obtained using a GeneChip WT PLUS Reagent kit (902280; Thermo Fisher Scientific) following the manufacturer's instructions. Hybridization on array strips was performed for 16 h at 48°C. The arrays were washed, stained, and scanned using a GeneAtlas HWS Kit (901667; Thermo Fisher Scientific) on the GeneAtlas Fluidics and Imaging Station. RNA-seq libraries were done from root total RNA using Standard RNA-Seq protocol method (PolyA selection for mRNA species) by the *Genwiz* company.

Nucleosome density

Root samples were frozen in liquid nitrogen and nuclei were isolated using Nuclei Isolation Buffer (NIB: PIPES-KOH pH 7.6 20 mM, Hexylene Glycol 1 M, $MgCl_2$ 10 mM, EGTA 1 mM, NaCl 20 mM, KCl 60 mM, Triton-x100 0.5%, β-mercaptoethanol 5 mM, Protease Inhibitor Cocktail (Roche)), filtrated with

100 µm and 40 µm cell strainers, and centrifuge for 5 minutes at 1500 G. Nuclei were resuspended in MNase reaction buffer (Tris-HCl, pH 8.0 20 mM, NaCl 5 mM, $CaCl_2$ 2.5 mM). Samples were divided into 2, and either treated with 5 uL of MNase (Takara, diluted 1/20) or mock-treated, for 7 minutes. Reactions were stopped with 2 µL of EDTA 0.5 mM, treated with RNAse A before DNA purification by phenol-chloroform and precipitation by ethanol. DNA was quantified using qPCR. Primers used for quantification are described in Supplemental dataset. Quantification of digested samples was normalized using non-digested samples and the ratio obtained for sequences located at the relative +1 position of the housekeeping gene *AP2M (AT5G46630*).

DNA methylation

DNA methylation was quantified by McrBC-qPCR, as described in (Martin *et al.*, 2009). DNA was extracted by DNeasy plant mini kit (QIAGEN), and each sample were subjected to two reactions with or without McrBC (NEB) for overnight digestion. The reaction was stopped by 20 min incubation at 65°C. DNA methylation quantification is realized by qPCR using control samples (not treated with enzyme). Percentage of methylated DNA was calculated as $100\text{-}2^{e(\text{-dCp*100})}$, where Cp is the threshold cycle; dCp is the difference of Cp between digested and mock samples. Primers used for quantification are described in Supplemental dataset.

Small RNA Northern blot (electrophoresis, labelling, hybridization, quantification)

Northern blot of small RNAs was performed as described in (Blevins, 2017) with minor modifications. Total RNA from each sample were loaded on denaturing polyacrylamide-urea gels and separated using 4 hours of migration. Small RNAs were transferred overnight on nylon membrane (Whatman® 0,45µm), and cross-linked with EDC (Thermo Scientific™ Pierce™). Membranes were pre-hybridized 4 hours with PerfectHyb Plus Hybridization Buffer (Sigma), and hybridized overnight with radioactive probes. Probes were generated from denatured PCR products labelled with ^{32}P-CTP by Klenow enzyme (Fig. 3 and Supplemental Fig. 3), or from single-strand oligonucleotides labelled with ^{32}P-ATP by T4PNK (Fig. 5). List of primers and oligonucleotides used for probe design is described in

Supplemental dataset. Hybridized membranes were washed with 2X SSC 0,1% SDS solution and exposed to CL-XPosure Film (ThermoFisher). Intensity of signal was quantified using ImageJ.

Run-on

Root samples were frozen in liquid nitrogen and nuclei were isolated using NIB described above. Nuclei were resuspended in 75 µL of 1.3X transcription buffer (65 mM Tris HCl pH 8, 6.5 mM MgCl2, 6.5 mM KCl), and mixed with 25 µL of a solution containing ATP, CTP, GTP (final concentration at 0.75 mM) and biotinylated UTP (Biotin-16-UTP, Roche, final concentration 0.25 mM). Transcription reactions were performed at 30°C for 40 min. Samples RNA were extracted using TRI REAGENT (*MRC*) and treated with DNase. Biotinylated RNAs were recovered using streptavidin-coated beads (M-280 Streptavidin Dynabeads, Invitrogen). Following retrotranscription (Superscript III, Invitrogen)), accumulation of nascent transcripts was measured by qRT-PCR (LightCycler 480, *Roche Diagnostics*) using the SYBR^{R} Premix Ex Taq^{TM} (*TaKaRa*). Gene expression was normalized using *UBQ10* as an internal standard.

Cell fractionation

Root samples were frozen in liquid nitrogen and nuclei were isolated using NIB as described above. Supernatant were recovered and taken as representative of the cytoplasmic fraction. Nuclei were further purified using a 30% Percoll cushion. RNAs were extracted from both fraction using LS Trizol (Invitrogen) and resuspended in 50% formamide.

sRNAs sequencing

Small RNA sequencing was performed from total RNA by the company *Fasteris*. Total RNA from each sample was size selected for 18 to 50-nucleotide sRNAs using denaturing polyacrylamide-urea gels. Sequencing was done on a NextSeq instrument using 50-base single-end mode.

Bioinformatics analysis

Transcriptomics analyses and identification of DEGs. Normalization of the microarray raw intensities was performed using the Robust Multiarray Averaging (RMA) method from the *oligo* R package of BioConductor. DEGs were identified using the limma R package. A gene was identified as a DEG in a selected genotype when the adjusted P-value was <0.05. *Analysis of RNA-seq data*. Initial paired-end RNA-Seq reads were pre-processed using *fastp.* High quality reads from *WT* and *clf-29 lhp1-4* were aligned to the reference genome (*A. thaliana* TAIR10 release) using *STAR.* To analyze mis-splicing of mRNA, STAR output bam files were converted to bed format and we selected the reads mapping on both (i) at least 75 bp of the exons and (ii) at most 10bp of their flanking introns, using the intersect function of the *bedtools* package. The indels & polymorphisms were detected from *STAR* output bam files with the *mpileup* function of *bcftools.* Normalization and visualization functions from *DIANE* (Cassan *et al.*, 2021) were used to normalize, and explore the counts from RNA-Seq data. *Venn diagram and intersection of gene lists.* Venn diagrams were performed using *VennDiagram* or *ggVennDiagram* R packages. Determination whether the overlap between different lists of genes is higher than randomly expected, and its value of significance, was tested and calculated using *GeneSect* (http://virtualplant.bio.nyu.edu). *Genesect* uses a bootstrapping-based method, consisting of selecting randomly 1,000 gene lists of Arabidopsis AGI having the same size as the observed lists, and counting the number of times that the intersection size of the random lists is equal or higher than the intersection observed for the two tested gene lists, reaching to a P-value (Krouk *et al.*, 2010). *Gene ontologies*. Gene ontology enrichment were interrogated using the *ClusterProfiler* R package, using the biological process function. The enrichment tests correspond to Fisher's exact tests, determining if the proportion of a specific ontology in a list of interest is statistically higher than expected in all Arabidopsis genes. P-values from those tests are corrected for multiple testing, and the threshold for over-representation tests was set to 0.05. In the graphical results, each significant ontology is thus characterized by an adjusted P-value, and a "Count", representing the number of genes associated to this ontology in the list of interest. *Small RNA sequencing analysis*.

Small RNA-seq was analyzed using the *Shortstack* package, after filtering the fastq files from 21-, 23, 24 and 33 nt reads. In order to analyze sRNAs associated with degradation of mRNA, cluster of sRNAs obtained by *Shortstack* were first filtered to only consider for the rest of the analysis those mapping to coding genes. We then applied (i) a coverage filter, with at least 64 reads associated to a coding gene using all the libraries, (ii) a strand specific filter, keeping only clusters of sRNAs for which sRNA strand corresponds to mRNA strand, and (iii) a filter for a minimum complexity at 0.7 provided by *Shortstack*. *Normalization of sRNA-seq read density using miRNAs expression*. In order to normalize the density of sRNA reads between WT and *clf-29 lhp1-4*, we used absolute values of several miRNAs expression measured by stem-loop quantitative PCR. This analysis reports that these miRNAs are not significantly deregulated in *clf-29 lhp1-4* (Supplemental Fig. 3). Then we relate these values to the number of reads observed for these miRNAs in the sRNA-seq for the 2 libraries of WT and *clf-29 lhp1-4*. Compiled together, this leads to a normalization factor of 8.44 that we applied to calculate normalized read density between WT and *clf-29 lhp1-4*. *Metagene profiling.* Metagene profiling was performed using *computeMatrix* function of *deepTools*, using the transcript location of each gene and bigwig files with normalized density of sRNAs for each genotype as inputs. Ten groups of genes were made according to their level of expression in WT, with 0-10 corresponding to the first 10% (less expressed) and 90-100 corresponding to the last 10% (most expressed). Results were plotted with *plotProfile* function of *deepTools*. *Calculation of RNA degradation score and significance on PcG target genes.* sRNA-seq reads corresponding to degradation products were selected via the 3 filters previously described on *Shortstack* output (coverage, strand, and complexity), and quantified for each gene. RNA-Seq counts in the same conditions were used to divide the number of degradation products of each gene, providing a degradation rate, or degradation score, relative to gene expression. This normalization is designed to compensate the fact that degradation increases with gene expression. At this step, an expression filter was set to an average of 10 counts per condition to remove genes with extremely low expression in the RNA-Seq data. The median of degradation scores among a list of genes of interest (i.e. PcG target genes) was used as the observed value in a bootstrap

hypothesis test. The null distribution of this test was simulated as the median degradation scores of 1000 lists of genes randomly sampled genome-wide, the size of the lists being equal to the size of the list of interest. This null distribution served to assess whether the observed median degradation score in marked genes was significantly higher than expected in random lists of genes. The P-values, defined as the number of random lists with a median degradation score above the observed degradation score divided by 1000, were 0 for all markings and both genotypes. The derived Z-scores, defined as the number of standard deviations between the observed score and the mean of the null distribution, had values between 6 and 11, depending on the marking and genotype. Bootstrapping results were plotted with *ggplot2* in R.

Results

In order to get new insights into the role of PcG proteins in the regulation of actively transcribed genes, we first measured the expression of *NRT2.1* under highly transcription-permissive conditions (low nitrate) in mutants for PcG-related factors, including LHP1 and BMI1 proteins. As we already observed, *NRT2.1* expression was slightly higher in *clf-29* mutant than in WT (Fig. 1A) (Bellegarde *et al.*, 2018). Similarly, we observed that the expression of *NRT2.1* was increased in *lhp1-4* mutant in comparison to WT, and also to a lesser extent in *bmi1* mutants, presumably due to strong redundancy between BMI1 proteins (Merini *et al.*, 2017) (Fig. 1A). We concluded that *NRT2.1* might be targeted by both PRC1 and PRC2. This confirmed what might be suggested by epigenome-wide analyses, where *NRT2.1* is found to be marked by both H3K27me3 and H2AUb, and is described as an LHP1 target (Veluchamy *et al.*, 2016; Zhou *et al.*, 2017a). Several reports showed that mutants for CLF or LHP1 display similar phenotypes, and have thus proposed a functional redundancy between these two PcG-proteins in the regulation of genes with low expression (Derkacheva *et al.*, 2013; Veluchamy *et al.*, 2016; Wang *et al.*, 2016; Zhou *et al.*, 2017b). Therefore, we wondered whether CLF and LHP1 could also have a certain degree of redundancy in the regulation of *NRT2.1* and other highly

expressed genes. Surprisingly, the effect of concurrent loss of function for *CLF* and *LHP1* has not been described yet in the literature. To do so, we generated a double *clf-29 lhp1-4* mutant. This double mutant yields very few to almost no seeds, and was thus maintained at the heterozygous state for *clf-29* mutation. However, *clf-29 lhp1-4* mutant does not display severe PcG phenotype like callus-like structures observed in *clf swn* or *bmia/b/c* mutants (Bratzel *et al.*, 2010; Chanvivattana *et al.*, 2004) (Supplemental Fig. 1A, B). In opposition, even though *clf-29 lhp1-4* mutant shows strong growth defects, it displays a relatively correct organogenesis including a root system, leaves and the development of flowers and siliques. Concerning the root system, the length of the primary root of *clf-29 lhp1-4* was not significantly different of the WT, but the lateral root density was significantly reduced in *clf-29 lhp1-4* in comparison to WT (Supplemental Fig. 1C). The root anatomy of *clf-29 lhp1-4* mutant appears properly organized, as revealed by propidium iodide staining and confocal microscopy (Supplemental Fig. 1D). Therefore, we measured *NRT2.1* expression level in the roots of *clf-29 lhp1-4* mutant. Surprisingly, *NRT2.1* expression in *clf-29 lhp1-4* was largely reduced in comparison to the WT line, in opposition to the increase in expression observed in the two single mutants (Fig. 1A, Supplemental Fig. 2). The same observation was done using the *ProNRT2.1:LUC* reporter line, that we introduced in the *clf-29*, *lhp1-4* and *clf-29 lhp1-4* mutants (Fig. 1B), strongly suggesting again the redundancy between CLF and LHP1 in the regulation of *NRT2.1* expression. In agreement with these observations, the root nitrate influx, depending mostly on the function of NRT2.1, was strongly diminished in *clf-29 lhp1-4* mutants (Fig. 1C).

To explore the extent of the combined effects of the *clf-29* and *lhp1-4* mutations on the regulation of transcriptionally active genes, we generated and compared the root transcriptome profiles of WT and the double mutant *clf-29 lhp1-4* under low nitrate condition. In agreement with *NRT2.1* expression in *clf-29 lhp1-4*, we found that other highly expressed genes involved in N nutrition and usually co-expressed with *NRT2.1*, like *NRT1.1*, *NIA1* and *NIR1* also display a significantly reduced expression in *clf-29 lhp1-4* (Fig. 2A). Genome-wide, concurrent loss of CLF and LHP1 functions leads to the differential expression of almost 3500 genes (adjusted P-value < 0.05), in agreement with the major

function of PcG protein in the regulation of gene expression (Supplemental Table 1). To understand the effect of *CLF* and *LHP1* concurrent mutations on transcriptionally active genes at a larger scale, we tested if deregulation of genes in *clf-29 lhp1-4* was linked to their level of expression in WT. We clearly observed that genes showing high levels of expression in WT, like *NRT2.1*, tend to be down-regulated in *clf-29 lhp1-4*, in opposition to genes showing low levels of expression which tend to be up-regulated (Fig. 2B). In order to question whether down-regulation of highly expressed genes in *clf-29 lhp1-4* correlates with the presence of H3K27me3 and LHP1 at their loci, we isolated among the highly expressed genes those significantly down-regulated in *clf-29 lhp1-4*, and marked by both H3K27me3 and LHP1. To do so, we used ChIP-seq data from two studies that previously described H3K27me3 and LHP1 target genes (Roudier *et al.*, 2011; Veluchamy *et al.*, 2016). Even if the nitrate conditions are not identical between those studies and ours, it has been demonstrated that H3K27me3 profile is globally very robust in Arabidopsis, and does not change according to nitrate provision (Bellegarde *et al.*, 2018; Roudier *et al.*, 2011). This leads to a relatively low number of 70 genes, as most transcriptionally active genes are under-represented in H3K27me3 and LHP1 marked-genes (Fig. 2C). However, this number was significantly higher than randomly expected, demonstrating that genes highly expressed and marked by H3K27me3 and LHP1 in WT tend to be down-regulated in *clf-29 lhp1-4* (Supplemental Fig. 2). Finally, we analyzed the function of these 70 *NRT2.1*-like genes by looking at gene ontology enrichments. Although we interrogated a relatively low number of genes, we significantly identified several functional categories that were over-represented. In particular, we observed that “response to nitrate”, “inorganic ion transmembrane transport” or “inorganic anion transport” were among the most significantly over-represented functional categories (Fig. 2D). This led us to the conclusion that PcG proteins might target a group of transcriptionally active genes associated to metabolism and nutrition, including nitrate transport, and that the concurrent absence of regulation exerted by CLF and LHP1 lead to their down-regulation.

Excessive expression in plants can lead to the production of aberrant transcripts, which in turn can trigger post-transcriptional regulation like RNAi, or the establishment of transcriptional gene silencing

(Christie *et al.*, 2011; Liu and Chen, 2016). Given the redundancy between *CLF* and *LHP1* and the very high expression of *NRT2.1* transcript level in *clf-29* and *lhp1-4* single mutants, we reasoned that the decrease of *NRT2.1* transcript level in *clf-29 lhp1-4* might be due to the occurrence of a silencing mechanism in this genetic background. To understand which mechanism could drive the decrease in *NRT2.1* transcript level in *clf-29 lhp1-4* double mutant, we analyzed several chromatin-based signatures of transcriptional gene silencing or post-transcriptional gene silencing at the *NRT2.1* locus. We first compared the chromatin architecture between WT and *clf-29 lhp1-4* plants, by profiling nucleosome density and positioning at the *NRT2.1* locus. We did not observe any significant difference in *NRT2.1* nucleosome density or positioning in *clf-29 lhp1-4* (Fig. 3A). We also quantitatively analyzed DNA methylation, which is a hallmark of transcriptional gene silencing, at the *NRT2.1* locus using McrBC digestion combined with quantitative PCR. Again, no difference could be found in the DNA methylation levels at *NRT2.1* in *clf-29 lhp1-4* when compared to WT (Fig. 3B). Then, in order to look for a possible post-transcriptional regulation directed to *NRT2.1* transcripts, we analyzed the small RNAs (sRNAs) profile associated to *NRT2.1* by performing sRNAs Northen blot in WT and *clf-29 lhp1-4*. Several RNA-directed regulatory processes exist in plants to target and degrade aberrant endogenous or exogenous RNAs, including messenger RNAs (mRNAs) surveillance by RNA quality control (RQC) mechanisms, regulation by microRNAs (miRNAs), or double stranded RNA-mediated post-transcriptional gene silencing (Martinez de Alba *et al.*, 2013). The sRNAs profile associated to *NRT2.1* in WT and *clf-29 lhp1-4* did not show the presence of 21-, 23- or 24-nt sRNAs, that are typical of double stranded RNA-mediated post-transcriptional gene silencing mechanisms. However, it revealed in *clf-29 lhp1-4* a massive amount of degradation products (Fig. 3C). The signature of *NRT2.1* degradation was very intense in *clf-29 lhp1-4*, and interestingly traces of such degradation, although to a much lesser extent, also appear in the WT line. We also observed that *NRT2.1* degradation was likely established in *clf-29* and *lhp1-4* simple mutants, although the amount of degradation was clearly lower than in *clf-29 lhp1-4* (Supplemental Fig. 3). Interestingly, the moderate levels of *NRT2.1* RNA degradation observed in *clf-29* and *lhp1-4* single mutants compared

to that of *clf-29 lhp1-4* double mutant appears to be in line with their redundancy observed at the level of the regulation of *NRT2.1* gene expression. To confirm that the observed reduction of *NRT2.1* transcript level in *clf-29 lhp1-4* was due to a post-transcriptional mechanism, we performed run-on experiments in WT and *clf-29 lhp1-4.* We used *SOC1*, which is a typical target of PcG proteins, as a positive control of run-on experiments, showing an increase in transcription rate in *clf-29 lhp1-4* in comparison to the WT (del Olmo *et al.*, 2016). We observed that *NRT2.1* transcription rate in *clf-29 lhp1-4* was also higher in *clf-29 lhp1-4* than in the WT line, confirming that the decrease in *NRT2.1* transcript level was not due to a diminution in transcription (Fig. 3D). Interestingly, we also observed that the increase in *NRT2.1* transcription rate in *clf-29 lhp1-4* was not in a higher range than the increase in *NRT2.1* transcript level in *clf-29* or *lhp1-4* single mutants (Fig. 1A), suggesting that the concurrent mutations of *CLF* and *LHP1* do not lead to a burst of *NRT2.1* expression. Altogether, these observations suggest that CLF and LHP1 might not only control the level of *NRT2.1* expression, but also the integrity or the quality of *NRT2.1* transcript, as a large part of *NRT2.1* mRNAs are subjected to degradation in *clf-29 lhp1-4* mutant, leading to a reduced transcript level compared to WT.

Aberrant or defective transcripts are extremely difficult to isolate, as they are instantly degraded to avoid deleterious effect. Therefore, to assess the genome-wide implication of PcG proteins in the regulation of transcript integrity of transcriptionally active genes in Arabidopsis, we focused on the final products of mRNA degradation. We thus analyzed the profile of sRNAs in WT and *clf-29 lhp1-4* by performing sequencing of sRNAs (sRNA-seq). First, we observed at the global level that the size distribution of sRNA populations was strikingly different between WT and *clf-29 lhp1-4*. Indeed, typical sRNA abundant populations in WT, constituted by 21-, 24- or 33-nt sRNAs, were apparently absent in *clf-29 lhp1-4* (Fig. 4A). In opposition, *clf-29 lhp1-4* mutant line shows a massive increase in non-canonical sRNAs ranging continuously from the lowest to the highest sizes of the sRNA-seq (Fig. 4A). In *clf-29 lhp1-4*, this population of sRNAs, likely associated with mRNA degradation, represented more than 75% of the sRNA population, in opposition to less than 45% in the WT (Fig 4B). Therefore, the apparent down-representation of 21-, 24- and 33-nt sRNAs in *clf-29 lhp1-4* could likely be due to

an increase in sRNAs of all other sizes that would affect the whole sequencing representation. To confirm this hypothesis, we performed stem-loop qRT-PCR with primers specific to four representative miRNAs in the genome. The result indicates that there is absolutely no decrease in the abundance of these miRNAs in *clf-29 lhp1-4* (Supplemental Fig. 4). Therefore, to perform a correct quantitative comparison of sRNA-seq results between *clf-29 lhp1-4* and WT, we used those four miRNAs to generate a normalization factor based on the effective quantification by stem-loop qRT-PCR and on the number of reads obtained in the sRNA-seq data (described in the Methods section). A first targeted analysis of sRNA-seq revealed that genes with an expression profile similar to *NRT2.1* in *clf-29 lhp1-4* also display a strong accumulation of sRNAs at their locus in this genotype (Fig. 4C). Moreover, nearly all of these non-canonical sRNAs were oriented in the sense strand of the transcripts, strongly reinforcing the fact that they originate from degraded mRNAs (Supplemental Fig. 5). To properly extract sRNA reads corresponding to RNA degradation, we discarded reads that had the size of canonical sRNAs (21-, 23-, 24- and 33-nt) and used filters for complexity (i.e a broad distribution over a gene coding sequence), read sense (corresponding to the coding strand of the transcript), and minimal coverage of sRNA clusters. Then, we inferred the abundance of sRNA reads corresponding to RNA degradation in WT and *clf-29 lhp1-4* at the genome-wide level. We observed an important increase of sRNA abundance over genes in *clf-29 lhp1-4*, with genes globally accumulating as much as eight times more sRNAs in *clf-29 lhp1-4* than in WT (Fig. 4D). This confirmed at the genome-wide level what we observed on *NRT2.1* and other co-expressed genes. Then we wanted to compare between WT and *clf-29 lhp1-4* the extent of degradation products according to the level of expression in WT. To do so, we performed a metagene analysis representing the density of sRNA reads associated with RNA degradation for subpopulations of genes ranked by their level of expression in WT. In agreement with our previous observations, we could see an important increase in the degradation profile of transcripts in the double mutant, in comparison to the WT. In the WT, only the 10% most expressed genes display a discernable level of RNA degradation. In opposition, we observed for *clf-29 lhp1-4* intense profile of RNA degradation, that were especially strong for the

most expressed genes (Fig. 4E). The highest gene expression level is, the stronger the effect of *clf-29 lhp1-4* mutations is and leads to sRNA accumulation (Fig. 4E). Then we could infer a list of more than 4000 genes that show an increase (at least 2-fold) in their amount of transcript degradation-related reads in *clf-29 lhp1-4*. Next, we wanted to test whether these genes are, like *NRT2.*1, overall highly expressed and misregulated in *clf-29 lhp1-4*. We therefore compared these genes with the list of deregulated genes in *clf-29 lhp1-4* and with the 10% most expressed genes in WT. This analysis showed a significant overlap between highly transcribed genes in WT and those degraded and deregulated in *clf-29 lhp1-4* (Fig. 4F). These data confirm that the loss of both CLF and LHP1 gives rise to massive transcript degradation on genes that are highly expressed, and that this degradation globally affects the expression level of these genes.

Degradation of transcripts is an important mechanism in eukaryotes that allows suppressing mRNAs with aberrant features that may ultimately compromise cellular functions. Given the excessive amount of transcript degradation related to transcriptionally active genes in *clf-29 lhp1-4*, we analyzed whether transcripts integrity might be affected following the loss of both CLF and LHP1 function. To do so, we analyzed RNA-seq data performed on WT and *clf-29 lhp1-4* lines, and looked at several features associated to premature degradation of mRNAs, including longer transcript, splicing fidelity, presence of small insertions, deletions or unexpected polymorphisms. This analysis shows, for the whole genome as for genes targeted by H3K27me3 or LHP1, that mRNAs from *clf-29 lhp1-4* did not show higher signatures of transcript defects compared to those of WT (Supplemental Table 2). This suggests that transcriptional fidelity was not significantly affected in *clf-29 lhp1-4*. In another hand, we asked whether the excessive production of sRNAs observed in *clf-29 lhp1-4* is linked to the chromatin environment mediated by PcG complexes. To do so, we compared the sRNAs abundance on PcG-targeted genes and PcG-non targeted genes in WT and *clf-29 lhp1-4*. We calculated for each gene of the genome a score of degradation, corresponding to the amount of sRNAs at each locus, normalized by the level of expression. The score of RNA degradation was compared between the list of H3K27me3- or/and LHP1-targeted genes and lists of randomly selected

genes. In WT, we observed strikingly that genes marked by H3K27me3, LHP1, or both, have a significantly higher score of RNA degradation than the rest of the genome (Fig. 5A). This, surprisingly, shows that PcG-targeted genes are generally prone to mRNAs degradation in comparison to the rest of the genome. In another hand, two important observations emerged from the corresponding analysis in *clf-29 lhp1-4*. First, the score of degradation was much higher than in the WT, confirming that degradation of mRNAs globally increases in *clf-29 lhp1-4.* However, we observed that the score of degradation increased for PcG-targeted genes but also for the rest of the genome, to a very similar extent (Fig. 5A). This strongly suggests that the increase of mRNAs degradation in *clf-29 lhp1-4* is not solely affecting PcG-targeted genes but the whole genome, and therefore means that accumulation of sRNAs is not directly mediated by the loss of PcG-related chromatin environment, but rather indirectly by another mechanism. Interestingly, when looking more carefully at the expression of genes annotated with the function "exoribonuclease activity and mRNA catabolic process", we found that some of these genes, such as *XRN4*, *XRN3* or *XRN2*, could be slightly induced in *clf-29 lhp1-4* (Fig. 5B). This suggests that the machinery of mRNAs degradation could be altered and boosted in *clf-29 lhp1-4*. However, this would also be an indirect explanation, as *XRN2, XRN3* and *XRN4* are not targeted by PcG chromatin marks (Roudier *et al.*, 2011). To test the origin of sRNAs associated to excessive RNA degradation in *clf-29 lhp1-4*, we performed sRNA Northern blots targeting *NRT2.1* transcript on nuclear and cytoplasmic compartments in WT and *clf-29 lhp1-4*. This analysis showed that *NRT2.1*-associated sRNAs accumulate in *clf-29 lhp1-4* in both nuclear and cytoplasmic compartments, with a stronger accumulation in the cytoplasm (Fig. 5C). These results suggest that sRNAs associated with increased RNA degradation in *clf-29 lhp1-4* certainly originate from the cytoplasm and the nuclei. Therefore, we concluded that excessive mRNA degradation caused by the loss of both CLF and LHP1 are presumably due to a misregulation of RNA degradation machinery, affecting the expression of genes with major physiological functions.

Discussion

In this work, we explored the role of Arabidopsis PcG group proteins in the regulation of transcriptionally active genes, in relation with mRNA degradation. Several other recent reports have illustrated that PcG functions that are not restricted to genes with low expression. In Arabidopsis, a subset of H2AUb-marked genes corresponds to transcriptionally active genes (Zhou et al. 2017a). These genes are associated with metabolic functions, but the impact of PcG proteins regulation on them has not been analyzed yet. PcG proteins have usually been described for their roles in keeping the expression level of developmental genes to a minimum, creating a so called facultative heterochromatin. However, it has already been shown that PcG proteins do not only target genes with low expression. In our data, the list of genes that are highly expressed and targeted by PcG proteins is not enriched in development related loci but in genes involved in metabolism. Indeed, active PcG targets are highly enriched in genes involved in the transport of diverse nutrients such as nitrate, amino acids, sulfate, phosphate, potassium and water. The main model gene used in this study, *NRT2.1*, is also involved in metabolism. Loss of expression of the gene encoding the main nitrate transporter in *clf-29 lhp1-4* reduces dramatically the nitrate influx capacity of the root system. Transcriptionally active PcG proteins targets have already been characterized for their enrichment in metabolic function (van den Boom *et al.*, 2016; Zhou *et al.*, 2017a). This suggests that PcG proteins are not only important to maintain organ identity by repressing homeotic genes but also to ensure the integrity of active transcripts involved in major metabolic functions. Our previous study also described that *NRT2.1* is constantly marked by H3K27me3 even under transcriptionally permissive conditions (Bellegarde *et al.*, 2018). Loss of CLF function leads to a reduction of H3K27me3 levels at the *NRT2.1* locus, increasing its mRNA levels without altering its expression pattern (Bellegarde *et al.*, 2018). Here, using *NRT2.1* as a model for active genes, we showed that concurrent loss of LHP1 and

CLF, which are members of PRC1 and PRC2 respectively, leads surprisingly to down-regulation of *NRT2.1* expression. Defects in *NRT2.1* expression could be due to morphological changes, that are known to be frequent in Polycomb mutants. However, several observations are not in favor of this possibility. Indeed, we observed that the root development and anatomy does not seem particularly disrupted in *clf-29 lhp1-4*. In addition, we previously demonstrated that the regulation mediated by CLF does not modify the tissue-specific expression of *NRT2.1* (Bellegarde et al., 2018). Finally, and most notably, we observed a striking association between the down-regulation of *NRT2.1* in *clf-29 lhp1-4* and the accumulation of sRNAs originating from the *NRT2.1* transcript. Additional experiments will be helpful to characterize the cellular and morphological defects in *clf-29 lhp1-4*, but our results strongly suggest that *NRT2.1* down-regulation is due to excessive RNA degradation that occurs in *clf-29 lhp1-4*. At the genome-wide level, loss of CLF and LHP1 function also leads to the down-regulation of numerous metabolism-related active genes. The analysis of sRNAs profile at *NRT2.1* and genome-wide revealed that down-regulation of expression of active genes following loss of LHP1 and CLF is associated to the accumulation of sRNAs, resulting from post-transcriptional degradation of mRNAs. This might suggest a link between PcG-mediated chromatin environment and RQC mechanisms. Indeed, a recent study in yeast showed that the LSM2-8 XRN2 associated complex, localized in the nucleus, is involved in the repression of gene expression applied by PcG proteins (Mattout et al. 2020). They could show that the LSM2-8 complex and XRN2 are specifically degrading transcripts coming from genes subjected to PcG repression to enhance and ensure their transcriptional silencing. In our study, we observed, whatever the genotype, that PcG-targeted genes are more prone to mRNA degradation than the rest of the genome. However, by differentiating PcG-targeted genes and PcG-non targeted genes, we showed that the mechanism for degradation of mRNAs in PcG mutants is likely indirect, as genes that are not marked by H3K27me3 and LHP1 are also subjected to a similar enhancement of mRNA degradation in *clf-29 lhp1-4*. This suggests that the global enhancement of mRNA degradation in *clf-29 lhp1-4* might be due to a deregulation of the expression of genes important for RQC mechanisms. In agreement, we observed that key genes involved in mRNA

degradation are induced in *clf-29 lhp1-4.* Yet, this link between PcG and the regulation of mRNAs degradation genes also appears indirect, as potential candidates, such as *XRN* genes, are not targeted by PcG-related chromatin marks. Further experiments should be undertaken in order to identify the link between PcG proteins and the homeostasis of mRNA degradation. In conclusion, our study highlights that transcriptionally active genes and especially those targeted by PcG are prone to mRNAs degradation, and that PcG function is also essential to maintain an accurate level of mRNAs degradation to ensure a correct expression level of genes.

Author Contribution

DS performed most of the experiments. OC and CF performed most of the bioinformatics analyses. FB and AG contributed to the initiation of the project. FB, GS and JB contributed to the generation and the first characterization of the double *clf-29 lhp1-4* mutant. JA and TL contributed to the analysis of sRNAs after cell fractionation. AM and DS conceived the project and designed the experiments. AM and DS interpreted the data with the help of every authors, and wrote the manuscript.

Data availability

Data generated within this paper are available on the ArrayExpress website under the accession numbers E-MTAB-9792, E-MTAB-9793 and E-MTAB-11263.

Conflict of Interest

The authors have no conflicts to declare

Funding

DS and FB were supported by a fellowship from the French Ministry of Higher Education and Research. OC were supported by a fellowship from the CNRS (Centre National de la Recherche Scientifique). AG and AM were supported by a fellowship from ANR (Agence Nationale de la Recherche) (ANR14-CE19-0008 IMANA).

References

Bellegarde F, Herbert L, Sere D, Caillieux E, Boucherez J, Fizames C, Roudier F, Gojon A, Martin A. 2018. Polycomb Repressive Complex 2 attenuates the very high expression of the Arabidopsis gene *NRT2.1*. Scientific Reports **8**, 7905.

Blevins T. 2017. Northern Blotting Techniques for Small RNAs. Methods in Molecular Biology **1456**, 141-162.

Bratzel F, Lopez-Torrejon G, Koch M, Del Pozo JC, Calonje M. 2010. Keeping cell identity in Arabidopsis requires PRC1 RING-finger homologs that catalyze H2A monoubiquitination. Current Biology **20**, 1853-1859.

Brien GL, Healy E, Jerman E, Conway E, Fadda E, O'Donovan D, Krivtsov AV, Rice AM, Kearney CJ, Flaus A, McDade SS, Martin SJ, McLysaght A, O'Connell DJ, Armstrong SA, Bracken AP. 2015. A chromatin-independent role of Polycomb-like 1 to stabilize p53 and promote cellular quiescence. Genes & Development **29**, 2231-2243.

Cassan O, Lebre S, Martin A. 2021. Inferring and analyzing gene regulatory networks from multi-factorial expression data: a complete and interactive suite. BMC Genomics **22**, 387.

Chanvivattana Y, Bishopp A, Schubert D, Stock C, Moon Y-H, Sung ZR, Goodrich J. 2004. Interaction of Polycomb-group proteins controlling flowering in Arabidopsis. Development **131**, 5263-5276.

Christie M, Brosnan CA, Rothnagel JA, Carroll BJ. 2011. RNA decay and RNA silencing in plants: competition or collaboration? Frontiers in Plant Sciences **2**, 99.

del Olmo I, López JA, Vázquez J, Raynaud C, Piñeiro M, Jarillo JA. 2016. Arabidopsis DNA polymerase ϵ recruits components of Polycomb repressor complex to mediate epigenetic gene silencing. Nucleic Acids Research **44**, 5597-5614.

Derkacheva M, Steinbach Y, Wildhaber T, Mozgova I, Mahrez W, Nanni P, Bischof S, Gruissem W, Hennig L. 2013. Arabidopsis MSI1 connects LHP1 to PRC2 complexes. The EMBO Journal.

Forderer A, Zhou Y, Turck F. 2016. The age of multiplexity: recruitment and interactions of Polycomb complexes in plants. Current Opinion in Plant Biology **29**, 169-178.

Girin T, El-Kafafi E-S, Widiez T, Erban A, Hubberten H-M, Kopka J, Hoefgen R, Gojon A, Lepetit M. 2010. Identification of Arabidopsis Mutants Impaired in the Systemic Regulation of Root Nitrate Uptake by the Nitrogen Status of the Plant. Plant Physiology **153**, 1250-1260.

Gonzalez I, Munita R, Agirre E, Dittmer TA, Gysling K, Misteli T, Luco RF. 2015. A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. Nature Structural and Molecular Biology **22**, 370-376.

Hecker A, Brand LH, Peter S, Simoncello N, Kilian J, Harter K, Gaudin V, Wanke D. 2015. The Arabidopsis GAGA-Binding Factor BASIC PENTACYSTEINE6 Recruits the POLYCOMB-REPRESSIVE COMPLEX1 Component LIKE HETEROCHROMATIN PROTEIN1 to GAGA DNA Motifs. Plant Physiology **168**, 1013-1024.

Jacquot A, Chaput V, Mauries A, Li Z, Tillard P, Fizames C, Bonillo P, Bellegarde F, Laugier E, Santoni V, Hem S, Martin A, Gojon A, Schulze W, Lejay L. 2020. NRT2.1 C-terminus phosphorylation prevents root high affinity nitrate uptake activity in Arabidopsis thaliana. New Phytologist.

Krouk G, Mirowski P, LeCun Y, Shasha DE, Coruzzi GM. 2010. Predictive network modeling of the high-resolution dynamic plant transcriptome in response to nitrate. Genome Biology **11**, R123.

Lecona E, Rojas LA, Bonasio R, Johnston A, Fernandez-Capetillo O, Reinberg D. 2013. Polycomb protein SCML2 regulates the cell cycle by binding and modulating CDK/CYCLIN/p21 complexes. PLoS Biology **11**, e1001737.

Livak, K. J. and Schmittgen T. D. 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods **25**(4): 402-408.

Li Z, Fu X, Wang Y, Liu R, He Y. 2018. Polycomb-mediated gene silencing by the BAH-EMF1 complex in plants. Nature Genetics **50**, 1254-1261.

Liu L, Chen X. 2016. RNA Quality Control as a Key to Suppressing RNA Silencing of Endogenous Genes in Plants. Molecular Plant **9**, 826-836.

Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, Morin H, Pitrat M, Dogimont C, Bendahmane A. 2009. A transposon-induced epigenetic change leads to sex determination in melon. Nature **461**, 1135-1138.

Martinez de Alba AE, Elvira-Matelot E, Vaucheret H. 2013. Gene silencing in plants: a diversity of pathways. Biochimica et Biophysica Acta **1829**, 1300-1308.

Merini W, Calonje M. 2015. PRC1 is taking the lead in PcG repression. The Plant Journal **83**, 110-120.

Merini W, Romero-Campero FJ, Gomez-Zambrano A, Zhou Y, Turck F, Calonje M. 2017. The Arabidopsis Polycomb Repressive Complex 1 (PRC1) Components AtBMI1A, B, and C Impact Gene Networks throughout All Stages of Plant Development. Plant Physiology **173**, 627-641.

Mohd-Sarip A, Lagarou A, Doyen CM, van der Knaap JA, Aslan Ü, Bezstarosti K, Yassin Y, Brock HW, Demmers JAA, Verrijzer CP. 2012. Transcription-Independent Function of Polycomb Group Protein PSC in Cell Cycle Control. Science **336**, 744-747.

Mozgova I, Hennig L. 2015. The polycomb group protein regulatory network. Annual Review in Plant Biology **66**, 269-296.

Mozgova I, Kohler C, Hennig L. 2015. Keeping the gate closed: functions of the polycomb repressive complex PRC2 in development. The Plant Journal **83**, 121-132.

Pherson M, Misulovin Z, Gause M, Mihindukulasuriya K, Swain A, Dorsett D. 2017. Polycomb repressive complex 1 modifies transcription of active genes. Science Advances **3**, e1700944.

Roudier F, Ahmed I, Berard C, Sarazin A, Mary-Huard T, Cortijo S, Bouyer D, Caillieux E, Duvernois-Berthet E, Al-Shikhley L, Giraut L, Despres B, Drevensek S, Barneche F, Derozier S, Brunaud V, Aubourg S, Schnittger A, Bowler C, Martin-Magniette M-L, Robin S, Caboche M, Colot V. 2011. Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. The EMBO Journal **30**, 1928-1938.

Schuettengruber B, Bourbon HM, Di Croce L, Cavalli G. 2017. Genome Regulation by Polycomb and

Trithorax: 70 Years and Counting. Cell **171**, 34-57.

Sequeira-Mendes J, Araguez I, Peiro R, Mendez-Giraldez R, Zhang X, Jacobsen SE, Bastolla U, Gutierrez C. 2014. The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States. The Plant Cell **26**, 2351-2366.

Takada S, Goto K. 2003. TERMINAL FLOWER2, an Arabidopsis Homolog of HETEROCHROMATIN PROTEIN1, Counteracts the Activation of FLOWERING LOCUS T by CONSTANS in the Vascular Tissues of Leaves to Regulate Flowering Time. The Plant Cell **15**, 2856-2865.

Turck F, Roudier F, Farrona S, Martin-Magniette ML, Guillaume E, Buisine N, Gagnot S, Martienssen RA, Coupland G, Colot V. 2007. Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. PLoS Genetics **3**, e86.

van den Boom V, Maat H, Geugien M, Rodriguez Lopez A, Sotoca AM, Jaques J, Brouwers-Vos AZ, Fusetti F, Groen RW, Yuan H, Martens AC, Stunnenberg HG, Vellenga E, Martens JH, Schuringa JJ. 2016. Non-canonical PRC1.1 Targets Active Genes Independent of H3K27me3 and Is Essential for Leukemogenesis. Cell Reports **14**, 332-346.

Varkonyi-Gasic E, Wu R, Wood M, Walton EF, Hellens RP. 2007. Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. Plant Methods **3**, 12.

Veluchamy A, Jegu T, Ariel F, Latrasse D, Mariappan KG, Kim SK, Crespi M, Hirt H, Bergounioux C, Raynaud C, Benhamed M. 2016. LHP1 Regulates H3K27me3 Spreading and Shapes the Three-Dimensional Conformation of the Arabidopsis Genome. PLoS ONE **11**, e0158936.

Wang H, Liu C, Cheng J, Liu J, Zhang L, He C, Shen WH, Jin H, Xu L, Zhang Y. 2016. Arabidopsis Flower and Embryo Developmental Genes are Repressed in Seedlings by Different Combinations of Polycomb Group Proteins in Association with Distinct Sets of Cis-regulatory Elements. PLoS Genet **12**, e1005771.

Wang Q, Shen WH. 2018. Chromatin modulation and gene regulation in plants: insight about PRC1 function. Biochemical Society Transactions **46**, 957-966.

Xu L, Shen WH. 2008. Polycomb silencing of KNOX genes confines shoot stem cell niches in Arabidopsis. Current Biology **18**, 1966-1971.

Zhou Y, Romero-Campero FJ, Gomez-Zambrano A, Turck F, Calonje M. 2017a. H2A monoubiquitination in Arabidopsis thaliana is generally independent of LHP1 and PRC2 activity. Genome Biology **18**, 69.

Zhou Y, Tergemina E, Cui H, Forderer A, Hartwig B, Velikkakam James G, Schneeberger K, Turck F. 2017b. Ctf4-related protein recruits LHP1-PRC2 to maintain H3K27me3 levels in dividing cells in Arabidopsis thaliana. Proceedings of the National Academy of Sciences, USA **114**, 4833-4838.

Figure Legends

Figure 1: Expression of transcriptionally active *NRT2.1* is controlled by PcG proteins. A. Expression of *NRT2.1* quantified by RT-qPCR. Relative expression levels were first calculated based on *UBQ10* as internal control and normalized to the expression level in WT. Data are presented as the mean (with SD) of at least 4 biological repeats. *P < 0.05, **P < 0,01, unpaired t-test. **B.** Expression of *LUC* in *ProNRT2.1:LUC* line quantified by RT-qPCR. Relative expression levels were first calculated based on *UBQ10* as internal control and normalized to the expression level in WT. Data are presented as the mean (with SD) of at least 4 biological repeats. *P < 0.05, unpaired t-test. **C.** Root NO_3^- influx measured at the external concentration of 0.2 mM $^{15}NO_3^-$. Data are presented as the mean of at least 7 replicates (with SD). ****P < 0.0001, unpaired t-test.

Figure 2: Transcriptomic behavior of active genes in *WT* and *clf-29 lhp1-4*. A. Expression of *NRT2.1, NRT1.1, NIA1 and NIR1* calculated from transcriptomic data. Data are presented as normalized expression values for the 3 biological repeats used to perform transcriptome profiles. **B.** Scatter plot representing the relationship between expression level in WT and expression fold change (FC) between *clf-29 lhp1-4* and WT for differentially expressed genes (Adjusted P-value < 0.05). **C.** Venn diagram representing the overlap between transcriptionally active genes (corresponding to the 10% most expressed) marked by H3K27me3 and LHP1, and transcriptionally active genes co-regulated with *NRT2.1* in *clf-29 lhp1-4.* Significant overlap is indicated by asterisks (*** P < 0.001, non-parametric randomization test). **D.** Gene ontology analysis performed on the set of *NRT2.1*-like genes in *clf-29 lhp1-4*. The enrichment tests correspond to Fisher's exact tests. P-values are corrected for multiple testing, and the threshold for over-representation tests was set to 0.05. Gene counts represent the number of genes associated to a given ontology in the list of interest.

Figure 3: *clf-29 lhp1*-4 mutations lead to the accumulation of *NRT2.1* transcript degradation products. A. Nucleosome density quantified by MNase-qPCR at the *NRT2.1* locus. Relative nucleosome density was first calculated relative to the input of total chromatin, and normalized by

the nucleosome density at the +1 nucleosome of *AP2M* (*AT5G46630*, encoding a member of the clathrin adaptor complexes). Data are presented as the mean (with SD) of 4 biological replicates. **B.** DNA methylation levels were quantified by McrBC-qPCR. Quantification by qRT-PCR is shown as the percentage of methylated DNA. *AP2M* (*AT5G46630*, encoding a member of the clathrin adaptor complexes) and *AT4TE09085* serve as negative and positive methylated controls, respectively. Data are presented as the mean (with SD) of 4 biological replicates. **C.** Northern blot analysis of small RNAs at the *NRT2.1* locus. Quantification of signal was performed by ImageJ from two replicates of independent Northern blots. **D.** Transcription rate of *NRT2.1* performed by run-on experiment and quantified by RT-qPCR. Relative expression levels were calculated based on *CLATHRIN* transcription rate as internal control. Data are presented as the mean (with SD) of at 3 biological repeats. *P < 0.05, unpaired t-test. Transcription rate of *SOC1* performed by run-on serves as positive control for the induction of expression in *clf-29 lhp1-4*, and was selected based on transcriptome analysis. **P < 0.01, unpaired t-test.

Figure 4: Small RNAs sequencing reveals massive accumulation of RNA degradation products in *clf-29 lhp1-4.* **A.** Size distribution of the reads obtained after sequencing of 2 small RNAs libraries from WT and *clf-29 lhp1-4*, expressed for each size (18- to 41-nt) in the percentage of the total amount of reads. Data are presented as the mean of the 2 libraries for each genotype. **B.** Pie chart of the read size composition of 2 small RNAs libraries from WT and *clf-29 lhp1-4*, showing the proportion of non-typical reads (all except 21-, 23-, 24-, and 33-nt) in the pool of small RNAs. Data are presented as the mean of the 2 libraries for each genotype. **C.** Normalized small RNAs density at *NRT2.1*, *NRT1.1* and *NAR2.1* loci in WT and *clf-29 lhp1-4*. **D.** Violin plots distribution showing the abundance of small RNAs associated with degradation over genes in WT and *clf-29 lhp1-4*. Data are presented as normalized read counts. **E.** Metagene analysis representing the normalized sRNA density in WT and *clf-29 lhp1-4* over 10 groups of genes ranked by their level of expression in WT. 0-10: 10% less expressed genes, 90-100: 10% most expressed genes. TSS, Transcriptional Start Site, TES, Termination Site. **F.** Venn diagram representing the overlap between genes transcriptionally active, genes with significant

increase in RNA degradation in *clf-29 lhp1-4*, and genes misregulated in *clf-29 lhp1-4* (P-value < 0.05, FC > 2). Significant overlap is indicated by asterisks (*** P < 0.001, non-parametric randomization test).

Figure 5: RNA degradation is enhanced for PcG target genes and in *clf-29 lhp1-4*. A. Score of RNA degradation in WT (top) and *clf29 lhp1-4* (bottom). Green single bar represents the median score of H3K27me3-marked genes, pink single bar represents the median score of LHP1-marked genes, blue single bar represents the median score for H3K27me3- and LHP1-marked genes. Green, pink and blue histograms represents the distribution of median scores for 1000 lists of randomly selected genes of the same size as H3K27me3-marked genes, LHP1-marked genes, and H3K27me3- and LHP1-marked genes, respectively. **B.** Heatmap of the expression of genes annotated in TAIR as involved in "exoribonuclease activity and mRNA catabolic process". Data are presented as normalized expression values for the 3 biological repeats used to perform transcriptome profiles. **C.** Northern blot analysis of small RNAs at the *NRT2.1* locus, preceded by cell fractionation of cytoplasmic and nuclear compartments. tRNA-VALINE and U6 serve as control for cytoplasmic and nuclear specificity, respectively. Quantification of *NRT2.1* signal was performed by ImageJ, and normalized to the signal calculated for tRNA-VALINE and U6. Cyt, cytoplasm, Nuc, nucleus, Val, tRNA-VALINE.

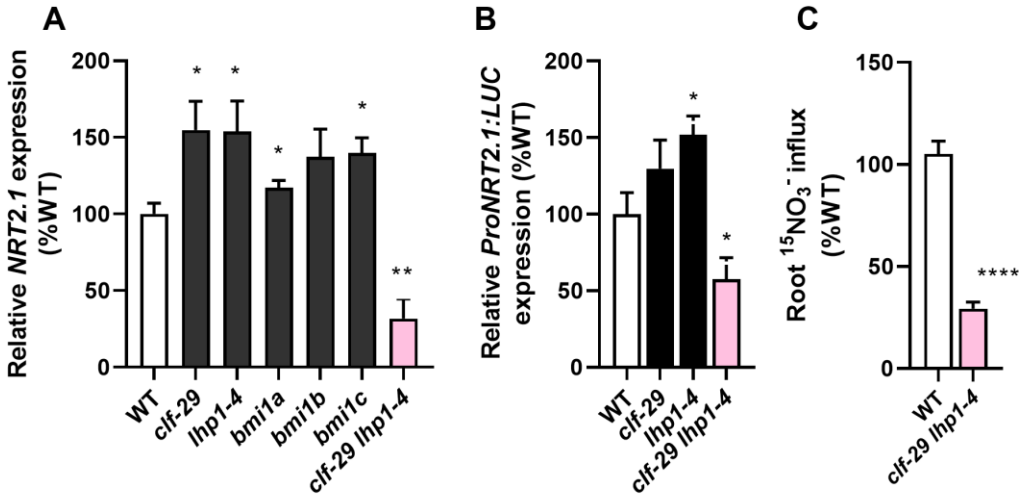

Figure 1: Expression of transcriptionally active NRT2.1 is controlled by PcG proteins. A. Expression of NRT2.1 quantified by RT-qPCR. Relative expression levels were first calculated based on UBQ10 as internal control and normalized to the expression level in WT. Data are presented as the mean (with SD) of at least 4 biological repeats. *P < 0.05, **P < 0,01, unpaired t-test. B. Expression of LUC in ProNRT2.1:LUC line quantified by RT-qPCR. Relative expression levels were first calculated based on UBQ10 as internal control and normalized to the expression level in WT. Data are presented as the mean (with SD) of at least 4 biological repeats. *P < 0.05, unpaired t-test. C. Root NO3- influx measured at the external concentration of 0.2 mM 15NO3-. Data are presented as the mean of at least 7 replicates (with SD). ****P < 0.0001, unpaired t-test.

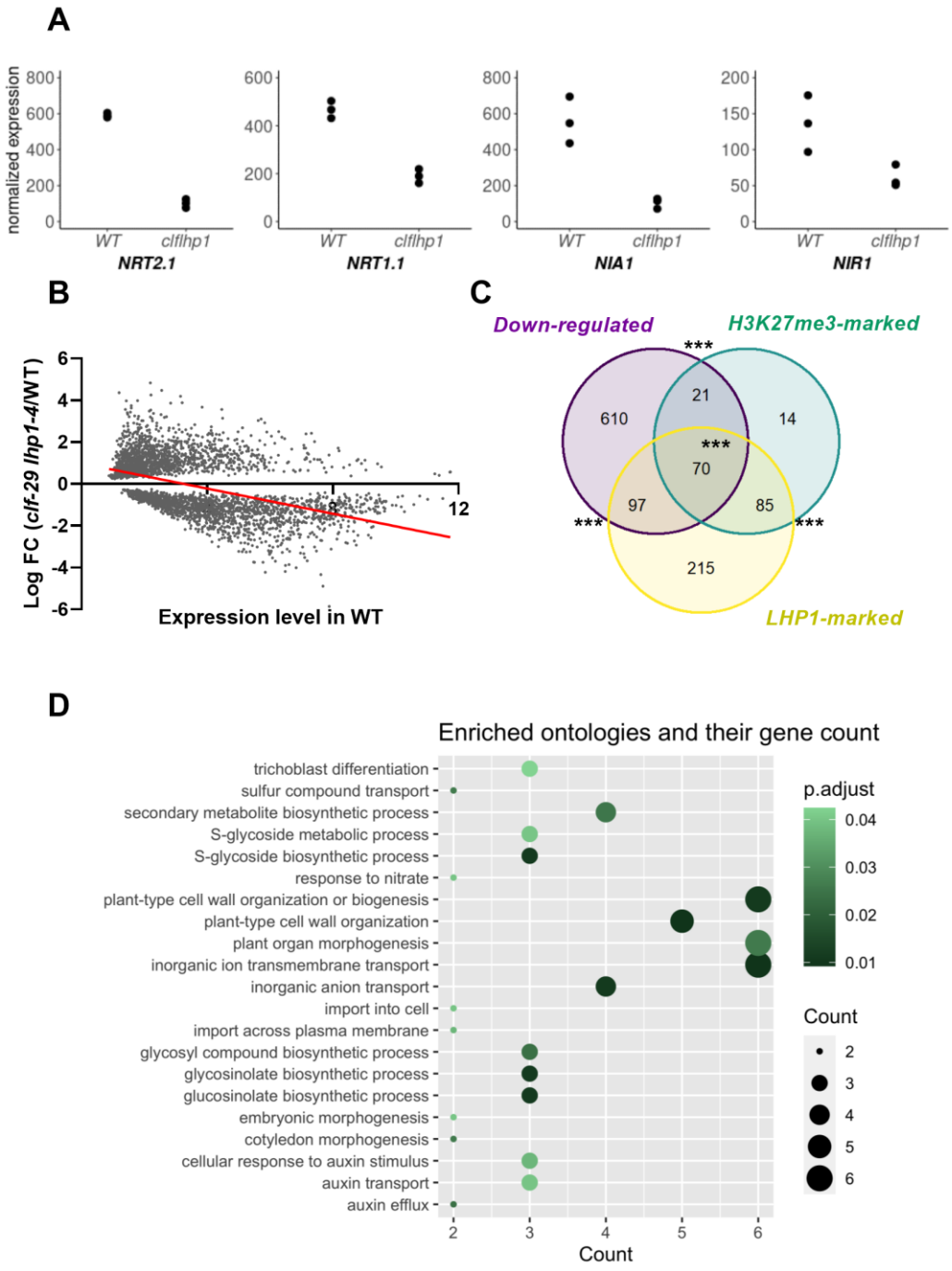

Figure 2: Transcriptomic behavior of active genes in WT and clf-29 lhp1-4. **A.** Expression of NRT2.1, NRT1.1, NIA1 and NIR1 calculated from transcriptomic data. Data are presented as normalized expression values for the 3 biological repeats used to perform transcriptome profiles. B. Scatter plot representing the relationship between expression level in WT and expression fold change (FC) between clf-29 lhp1-4 and WT for differentially expressed genes (Adjusted P-value < 0.05). C. Venn diagram representing the overlap between transcriptionally active genes (corresponding to the 10% most expressed) marked by H3K27me3 and LHP1, and transcriptionally active genes co-regulated with NRT2.1 in clf-29 lhp1-4. Significant overlap is indicated by asterisks (*** P < 0.001). D. Gene ontology analysis performed on the set of NRT2.1-like genes in clf-29 lhp1-4.

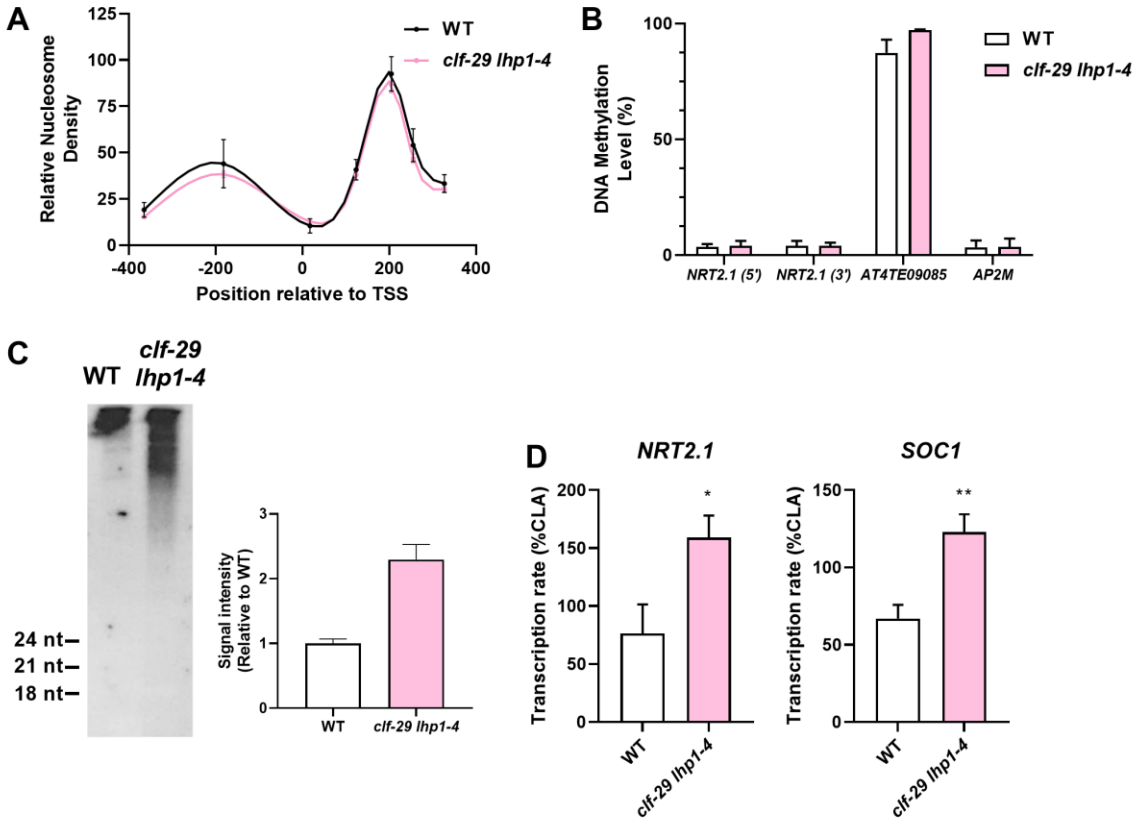

Figure 3: clf-29 lhp1-4 mutations lead to the accumulation of NRT2.1 transcript degradation products. A. Nucleosome density quantified by MNase-qPCR at the NRT2.1 locus. Relative nucleosome density was first calculated relative to the input of total chromatin, and normalized by the nucleosome density at the +1 nucleosome of AP2M (AT5G46630, encoding a member of the clathrin adaptor complexes). Data are presented as the mean (with SD) of 4 biological replicates. B. DNA methylation levels were quantified by McrBC-qPCR. Quantification by qRT-PCR is shown as the percentage of methylated DNA. AP2M (AT5G46630, encoding a member of the clathrin adaptor complexes) and AT4TE09085 serve as negative and positive methylated controls, respectively. Data are presented as the mean (with SD) of 4 biological replicates. C. Northern blot analysis of small RNAs at the NRT2.1 locus. Quantification of signal was performed by ImageJ from two replicates of independent Northern blots. D. Transcription rate of NRT2.1 performed by run-on experiment and quantified by RT-qPCR. Relative expression levels were calculated based on CLATHRIN transcription rate as internal control. Data are presented as the mean (with SD) of at 3 biological repeats. *P < 0.05, unpaired t-test. Transcription rate of SOC1 performed by run-on serves as positive control for the induction of expression in clf-29 lhp1-4, and was selected based on transcriptome analysis. **P < 0.01, unpaired t-test.

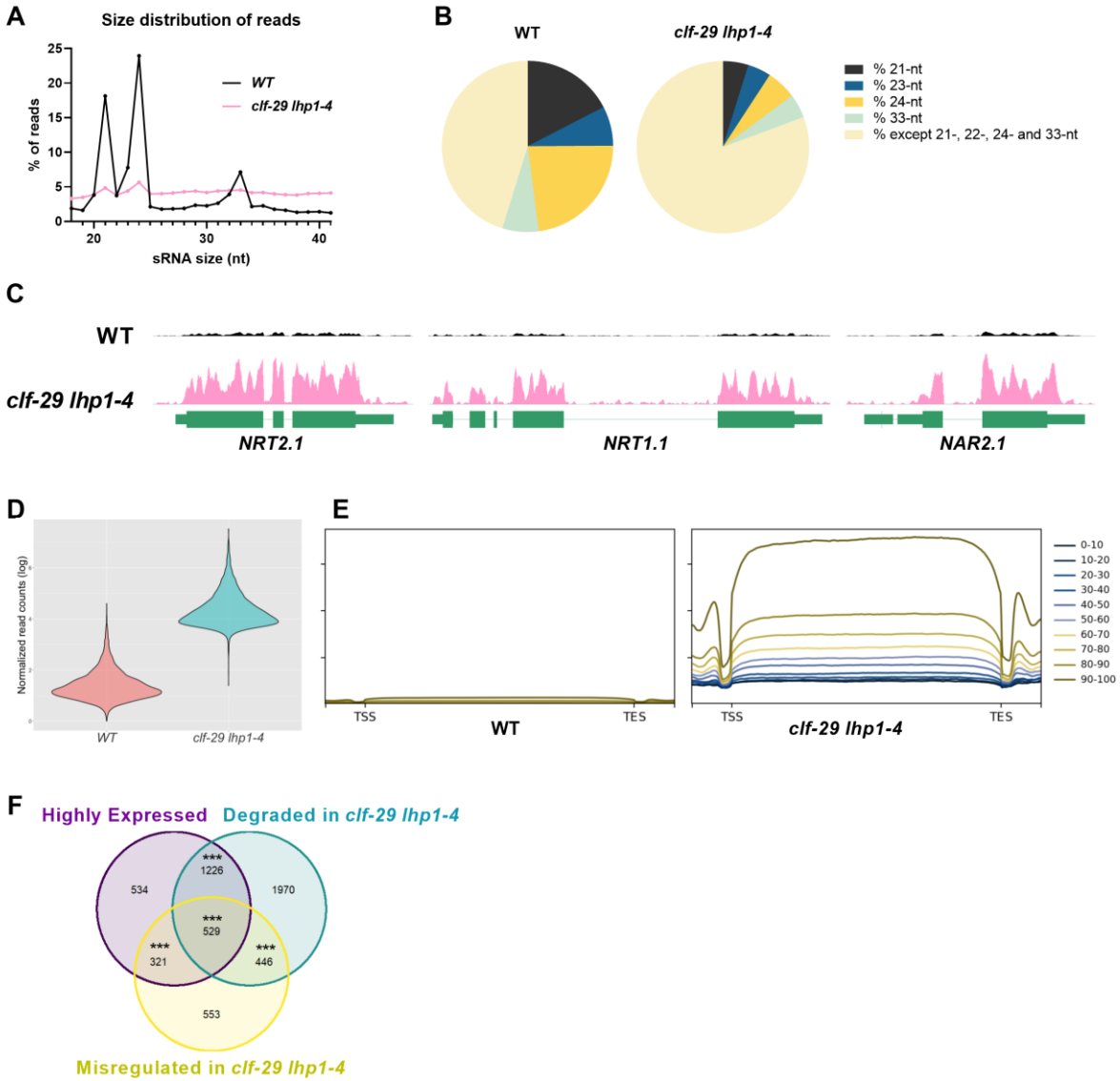

Figure 4: Small RNAs sequencing reveals massive accumulation of RNA degradation products in clf-29 lhp1-4. A. Size distribution of the reads obtained after sequencing of 2 small RNAs libraries from WT and clf-29 lhp1-4, expressed for each size (18- to 41-nt) in the percentage of the total amount of reads. Data are presented as the mean of the 2 libraries for each genotype. B. Pie chart of the read size composition of 2 small RNAs libraries from WT and clf-29 lhp1-4, showing the proportion of non-typical reads (all except 21-, 23-, 24-, and 33-nt) in the pool of small RNAs. Data are presented as the mean of the 2 libraries for each genotype. C. Normalized small RNAs density at NRT2.1, NRT1.1 and NAR2.1 loci in WT and clf-29 lhp1-4. D. Violin plots distribution showing the abundance of small RNAs associated with degradation over genes in WT and clf-29 lhp1-4. Data are presented as normalized read counts. E. Metagene analysis representing the normalized sRNA density in WT and clf-29 lhp1-4 over 10 groups of genes ranked by their level of expression in WT. 0-10: 10% less expressed genes, 90-100: 10% most expressed genes. TSS, Transcriptional Start Site, TES, Termination Site. F. Venn diagram representing the overlap between genes transcriptionally active, genes with significant increase in RNA degradation in clf-29 lhp1-4, and genes misregulated in clf-29 lhp1-4 (P-value < 0.05, FC > 2). Significant overlap is indicated by asterisks (*** P < 0.001).

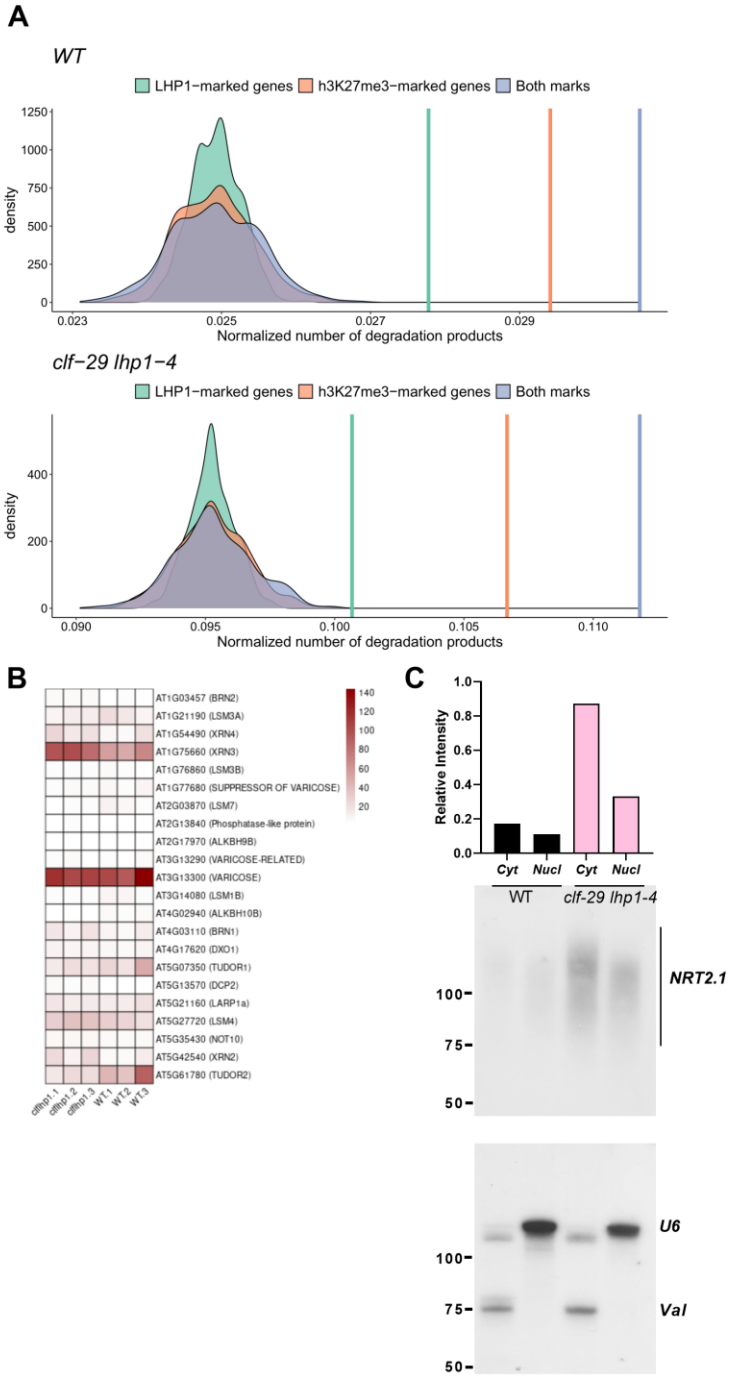

Figure 5: RNA degradation is enhanced for PcG target genes and in clf-29 lhp1-4. A. Score of RNA degradation in WT (top) and clf29 lhp1-4 (bottom). Green single bar represents the median score of H3K27me3-marked genes, pink single bar represents the median score of LHP1-marked genes, blue single bar represents the median score for H3K27me3- and LHP1-marked genes. Green, pink and blue histograms represents the distribution of median scores for 1000 lists of randomly selected genes of the same size as H3K27me3-marked genes, LHP1-marked genes, and H3K27me3- and LHP1-marked genes, respectively. B. Heatmap of the expression of genes annotated in TAIR as involved in "exoribonuclease activity and mRNA catabolic process". Data are presented as normalized expression values for the 3 biological repeats used to perform transcriptome profiles. C. Northern blot analysis of small RNAs at the NRT2.1 locus, preceded by cell fractionation of cytoplasmic and nuclear compartments. tRNA-VALINE and U6 serve as control for cytoplasmic and nuclear specificity, respectively. Quantification of NRT2.1 signal was performed by ImageJ, and normalized to the signal calculated for tRNA-VALINE and U6. Cyt, cytoplasm, Nuc, nucleus, Val, tRNA-VALINE.