Alternative silencing states of Transposable Elements in Arabidopsis associated with H3K27me3

Valentin Hure, Florence Piron-Prunier, Tamara Yehouessi, Clémentine Vitte, Aleksandra Kornienko, Gabrielle Adam, Magnus Nordborg, Angélique Déléris

▶ **To cite this version:**

Valentin Hure, Florence Piron-Prunier, Tamara Yehouessi, Clémentine Vitte, Aleksandra Kornienko, et al.. Alternative silencing states of Transposable Elements in Arabidopsis associated with H3K27me3. Genome Biology, 2024, 26, pp.11. 10.1186/s13059-024-03466-6 . hal-04791390

HAL Id: hal-04791390
https://hal.science/hal-04791390v1

Submitted on 28 Jan 2025

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

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

Distributed under a Creative Commons Attribution 4.0 International License

Genome Biology

RESEARCH

Open Access

Alternative silencing states of transposable elements in Arabidopsis associated with H3K27me3

Valentin Hure[1†], Florence Piron-Prunier[1†], Tamara Yehouessi[1†], Clémentine Vitte[2], Aleksandra E. Kornienko[3], Gabrielle Adam[4], Magnus Nordborg[3] and Angélique Déléris[1*]

[†]Valentin Hure, Florence Piron-Prunier, and Tamara Yehouessi contributed equally to this work.

*Correspondence: angelique.deleris@i2bc.paris-saclay.fr

[1] Institute for Integrative Biology of the Cell (I2BC), Université Paris-Saclay, Centre National de La Recherche Scientifique (CNRS), Commissariat À L'EnergieAtomique (CEA), Gif-Sur-Yvette 91190, France
[2] Université Paris-Saclay, Institut National de la Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE), CNRS, AgroParisTech, Génétique Quantitative et Evolution (GQE), Gif-Sur-Yvette 91190, France
[3] Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna BioCenter (VBC), Vienna, Austria
[4] Institute of Plant Sciences Paris-Saclay (IPS2), Université Paris-Saclay, CNRS, INRAE, Gif-Sur-Yvette 91190, France

Abstract

Background: The DNA/H3K9 methylation and Polycomb-group proteins (PcG)-H3K27me3 silencing pathways have long been considered mutually exclusive and specific to transposable elements (TEs) and genes, respectively in mammals, plants, and fungi. However, H3K27me3 can be recruited to many TEs in the absence of DNA/H3K9 methylation machinery and sometimes also co-occur with DNA methylation.

Results: In this study, we show that TEs can also be solely targeted and silenced by H3K27me3 in wild-type Arabidopsis plants. These H3K27me3-marked TEs not only comprise degenerate relics but also seemingly intact copies that display the epigenetic features of responsive PcG target genes as well as an active H3K27me3 regulation. We also show that H3K27me3 can be deposited on newly inserted transgenic TE sequences in a TE-specific manner indicating that silencing is determined in *cis*. Finally, a comparison of Arabidopsis natural accessions reveals the existence of a category of TEs—which we refer to as "bifrons"—that are marked by DNA methylation or H3K27me3 depending on the accession. This variation can be linked to intrinsic TE features and to trans-acting factors and reveals a change in epigenetic status across the TE lifespan.

Conclusions: Our study sheds light on an alternative mode of TE silencing associated with H3K27me3 instead of DNA methylation in flowering plants. It also suggests dynamic switching between the two epigenetic marks at the species level, a new paradigm that might extend to other multicellular eukaryotes.

Keywords: Transposable elements (TE), Polycomb (PcG), DNA methylation, Epigenetics, Epigenomics, Natural variation

Background

Transposable elements (TEs) are repeated sequences that can potentially move and multiply in genomes. Transposition is typically deleterious and is tightly controlled by host organisms. On the other hand, TE mobilization can be an important factor in evolution

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

and adaptation, in particular by providing nearby genes with genetic or epigenetic regulatory modules that can impact transcriptional programs [1]. Amplification of a TE typically leads to the production of groups of copies that are identical in sequence upon insertion, thus producing so-called 'TE families'. Over time, TE insertions accumulate SNPs and indels, leading to the generation of mutated and truncated copies. These are often non-autonomous, either due to their lack of functional protein-coding genes, structural regions necessary for their life cycle, or both. In the first case, copies can nevertheless be mobilized using proteins produced in *trans* by copies of the same family. In the absence of new insertions, eventually, only highly degenerate relics with no potential for mobility will remain [2]; they will also disappear unless selectively maintained for their *cis* (or *trans*) regulatory impact [3].

A hallmark of TE silencing in many organisms is high DNA methylation at cytosines (5-methylcytosines). In plants, TE methylation can be found in three sequence contexts (CG, CHG, and CHH, where H can be any base except G), and is coupled with histone H3K9 dimethylation (H3K9me2). The combination of these two epigenetic marks prevents TE expression and transposition, thus guaranteeing genome integrity [4, 5]. DNA methylation is established at TEs upon their insertion in the genome by the RNA-directed DNA methylation pathway: TE-derived small RNAs (or small RNAs preexisting in the nucleus) guide methyltransferase DRM2 to homologous TE sequences to establish DNA methylation in all three sequences cytosine contexts [6–8]. Four different pathways amplify the established DNA methylation patterns and maintain them stably over time. CG, CHG, and CHH DNA methylation patterns are typically maintained by MET1 [9], CMT3 [10], and CMT2 and RdDM [11], respectively. While very stable across generations, DNA methylation can be removed by DNA glycosylases of the DEMETER-like family (DML): these constitutively prune the DNA methylation patterns [12], prevent spreading outside of primary targeted sequences [13] or actively participate in gene control, either at specific stages of development or by preventing stress-responsive genes from being locked in a constitutively silent state.

DNA methylation and H3K9me2 patterns are known to be stable throughout development and generations and are hallmarks of the so-called constitutive heterochromatin. On the other hand, H3K27me3 is associated with facultative heterochromatin and is a hallmark of transcriptional repression associated with protein- and miRNA-coding genes involved in development, reproduction, or metabolic and stress responses [14–17]. The presence of H3K27me3 is tightly regulated by a balance between active addition by Polycomb Repressive Complex 2 (PRC2) [18] and active removal by histone REF6-ELF6-JMJ13/30/32 demethylases of the KDM4/JMJD2 family. H3K27me3 deposition establishes cell identity and silencing of stress-responsive/metabolic genes and its removal can allow gene activation [19], contribute to shaping the spatial and temporal patterns of H3K27me3, reset the mark during reproduction [20, 21], and maintain H3K27me1 patterns [22].

Because of their different targets and contrasted properties, PcG-H3K27me3 and DNA methylation have been generally considered as mutually exclusive systems, specialized for the transcriptional silencing of genes and TEs, respectively. However, this dichotomous model has been challenged in the past decade. In both plants and mammals, targeting of PcG to TEs was initially evidenced by gain of H3K27me3 at many TE

sequences upon their loss of DNA methylation [23–29] in mutants unable to maintain DNA and H3K9 methylation, as well as in specific cell types where TEs are naturally demethylated [25, 30]. This implies that PcG could serve as a backup silencing system for hypomethylated TEs (compensation), and functional redundancy of the two systems has been indeed revealed at some TEs in both mammals and plants [3, 30, 31]. These observations also have two mechanistic implications: that PcG can be recruited to TEs, and that DNA methylation can exclude H3K27me3 deposition. Yet, in some instances, the two marks can co-occur [32–35] and cooperate in restricting TE activation upon biotic stress [3]. More recently, it was also reported that PcG, instead of the DNA methylation machinery, targets and represses TE in unicellular eukaryotes and plants from early lineages, thus illustrating an ancient role of PcG in regulating TE silencing [36, 37].

TE targeting by PcG has been well described in DNA methylation mutants of *A. thaliana*, but whether this can also occur in the presence of an active DNA methylation machinery, replacing DNA methylation, remains unclear. Moreover, the modalities of PcG recruitment at TEs remain to be fully elucidated: while some TEs seem to be marked by H3K27me3 as the result of neighboring gene spreading, observation of H3K27me3 mark at TEs located away from any H3K27me3 marked gene indicates possible specific *cis*-recruitment, but this has not been formally tested. Besides, recruitment of PcG to TEs in the absence of DNA methylation points to a possible de novo recruitment on newly inserted TE sequences but this is unknown and needs to be investigated.

In this study, we report that many TEs (>4000) are covered by the H3K27me3 histone mark in the *A. thaliana* genome in WT plants. These elements comprise not only short TE relics, but also intact copies, which are expected to be targets of DNA methylation. We show that these long copies present similarities to classical PcG genic targets: they display the chromatin hallmarks of genes responsive to environmental or developmental cues and their H3K27me3 patterns are actively regulated by histone H3K27 demethylases. Using transgenic constructs, we demonstrate that newly inserted TE sequences are able to recruit H3K27me3 de novo, with patterns specific to each copy, indicating that recruitment of H3K27me3 relies on intrinsic features of the TE copies themselves. Finally, natural variation between Arabidopsis accessions at orthologous TE copies reveals an epigenetic switch between the two silencing modes. Using a GWAS approach, we provide evidence that this epigenetic switch is linked to both *cis*- and *trans*-determinants.

Our results establish an alternative, *cis*-determined epigenetic control of TEs by PcG in wildtype *A. thaliana*; they further provide evidence that a given TE can change its epigenetic silencing state throughout its lifespan. This challenges the dogma that Arabidopsis TE are solely targeted by DNA methylation in nature.

Results

Many TEs are marked solely by H3K27me3 in wild-type Arabidopsis plants

When examining H3K27me3 ChIP-seq profiles generated for natural Arabidopsis accessions, we noticed that several TEs showed an absence of DNA methylation and were instead marked by H3K27me3 (Fig. 1A, Additional file 1: Fig. S1A). To determine their abundance in the genome, we defined four classes of TEs based on the length coverage of H3K27me3 marks and CG methylation, using these datasets and previously published

methylomes [38] (Fig. 1B, Additional file 1: Fig. S1B–C, Additional file 2). The "class 1" represents the majority of TEs, which are DNA methylated only with high levels of DNA methylation in all three sequence contexts. The "class 2" contains TEs that are both DNA methylated and marked by H3K27me3. These TEs were characterized by lower CG and CHG methylation levels than "class I" TEs in line with observations in mammals whereby CG methylation density constrains deposition of H3K27me3 [32, 33] and more CHH methylation (Additional file 1: Fig. S1C). The class which consists of TEs only marked by H3K27me3 ("class 3") represents 15% of the total TE number but only 8.4% of TE bases. These percentages indicate a bias in the size of the TEs targeted H3K27me3 (Fig. 1B versus Additional file 1: Fig. S1D). Accordingly, TE with H3K27me3 tends to be more present within short TE categories (particularly those with a size below 1 kb), while long TEs (with a size over 3.5 kb) are generally associated with DNA methylation (Fig. 1C)—although a small fraction of long TEs (> 3.5 kb) are marked by H3K27me3 alone (Fig. 1C, Additional file 1: Fig. S1E). Interestingly, inspection of their predicted protein-coding domains showed that these TEs contain a number of functional domains similar to that of mobile TEs (GAG, group-specific antigens or Capsid protein; PR, protease; INT, integrase; RT, reverse-transcriptase; RH, ribonuclease H) (Fig. 1D, Additional file 1: Fig. S1F). Thus, TEs which are expected targets of DNA methylation can also be marked with H3K27me3.

Interestingly, H3K27me3-marked long TEs (> 3.5 kb) are dispersed throughout the euchromatic arms and pericentromeric regions (Fig. 1A, E and Additional file 1: Fig. S1G), in contrast to the DNA-methylated ones, which are generally clustered in pericentromeres. This suggests that genomic location is not the major determinant of PcG targeting. Further analysis of H3K27me3-marked TEs showed that they are enriched in Helitrons for all size categories, and particularly enriched in Helitrons and Copia LTR retrotransposons for long copies (Fig. 1F). Helitrons are frequently found to contain transcription factors binding sites and can act as gene promoters/cis-regulatory modules [40–42]. Interestingly, the H3K27me3-marked TEs (class 3) are closer to genes than are DNA methylated TEs (Additional file 1: Fig. S1H) as also observed recently [37], in

(See figure on next page.)

Fig. 1 Many TEs are marked by H3K27me3 in wild-type Arabidopsis (Col-0). **A**. Representative genome browser view of H3K27me3 (red), DNA methylation (gray), and constitutive heterochromatin histone variant H2A.W (gray) levels in wild-type Col-0 for two TEs marked by H3K27me3 in WT. Gray, black, and blue bars represent TEs, TE-genes, and Protein Coding Genes annotations, respectively. **B**. Venn diagram showing the classes of TEs based on their epigenetic modifications. A TE was defined as "DNA methylated" (classes 1–2) when CG methylation was > 50% in average across TE length (with the rationale that CG and non-CG DNA methylation are coupled at TEs, *Du and Johnson, 2015*); to validate this choice, associated DNA methylation in non-CG context was analyzed concomitantly on the classes defined as shown in Fig. S1B. A TE was defined to be in a H3K27me3 state when a H3K27me3 peak was present on at least 20% of the TE length. **C**. Stacked bar plot showing the distribution of TE classes according to TE length. **D**. Box plot showing the number of predicted domains for LTR Copia among TE > 3.5 kb using DANTE [39], ns = not significant. **E**. Schematic chromosomes showing distribution of TE > 3.5 kb along genome. **F**. Stacked bar plot showing distribution of TE super-families according to TE length among all TEs or H3K27me3 marked TEs. Statistical analysis was performed using the hypergeometric test. * = *p*-value < 0.05; ** = *p*-value < 0.01; *** = *p*-value < 0.001; **** = *p*-value < 0.0001. **G**. Plot showing a correlation between the percentage of members marked by H3K27me3 and the number of TE copies per subfamily. Each dot is a TE family (< 300 members). **H**. Heatmaps based on H3K27me3 levels along each TE copy within one given TE family. Top panel: *ATCOPIA12* and bottom panel: *ATCOPIA48*

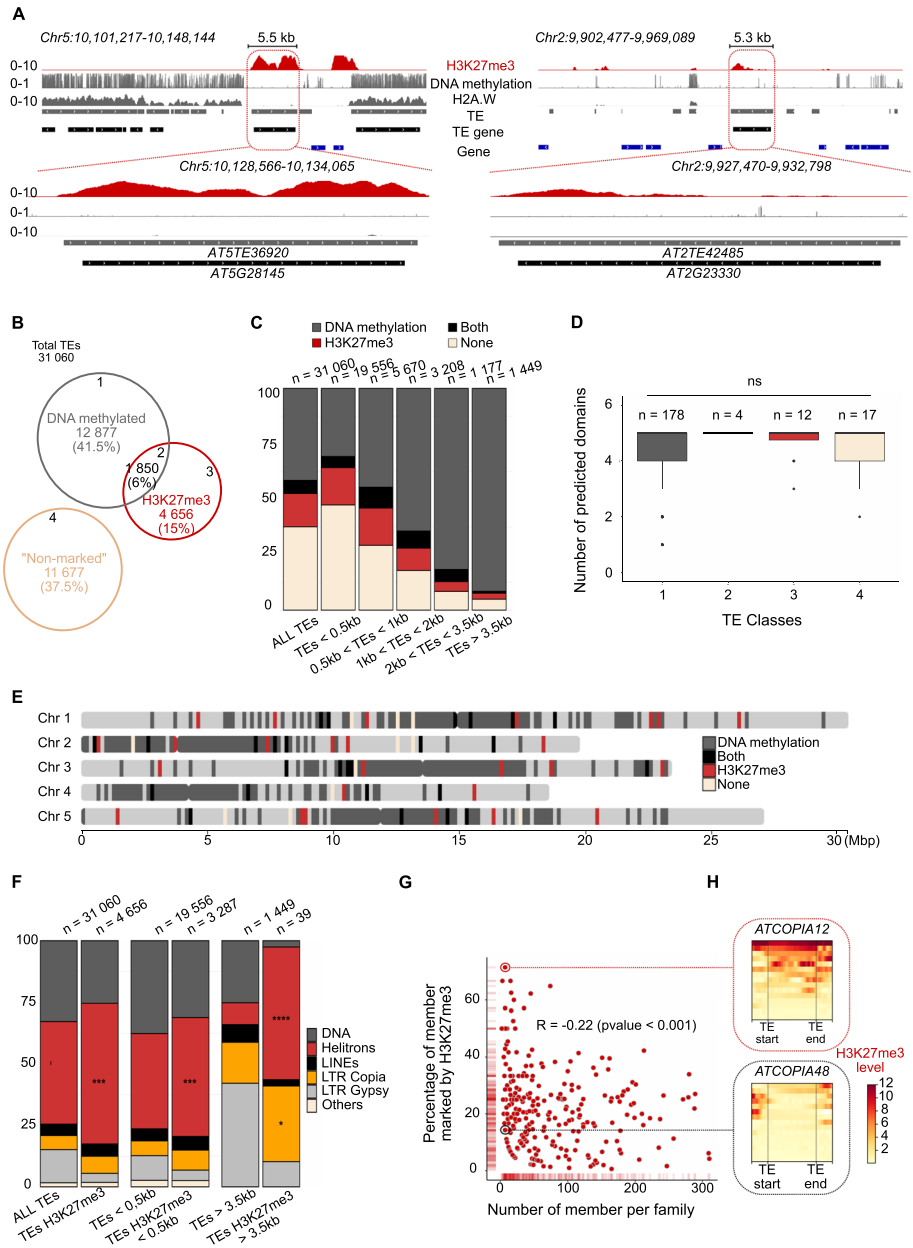

Fig. 1 (See legend on previous page.)

particular in the 500-bp upstream of promoters; H3K27me3-marked TEs in this interval comprise mostly Helitrons which suggests a role of these TEs as cis-regulatory modules, possibly with a dual activation/silencing function.

Comparison of epigenomic features (H3K27me3 and H3K9me2) at copies from the same family revealed intrafamily epigenetic variation (Additional file 1: Fig. S1H), with some TE copies within a family being marked by H3K27me3 while others are not. We further observed a weak but significant anti-correlation between copy number and percent of copies marked with H3K27me3 (Fig. 1G) suggesting that small TE families are more prone to be marked by H3K27me3 than families with large copy numbers. Nevertheless, small family size cannot fully explain epigenomic preference, as families

of similar copy numbers can exhibit very different H3K27me3 levels (exemplified by *ATCOPIA12* and *ATCOPIA48* families, Fig. 1H). Thus, while the TE characteristics tested could influence epigenetic status, they are not sufficient to explain it, suggesting that additional intrinsic features of TE copies participate in H3K27me3 recruitment.

H3K27me3-marked TEs have the epigenetic hallmarks of silent yet responsive genes and are transcriptionally repressed by PcG

H3K27me3 was previously shown to colocalize with other chromatin marks that can be associated with a dynamic transcriptional repression at protein-coding genes [43]. These include H2A.Z [44–47], a histone variant associated with responsive genes and transcriptional repression when ubiquitinated on Lysine 129 [48] as well as ubiquitination of H2A/Z (H2A/Zub) [49, 50], which is deposited by the Polycomb repressive complex PRC1 [51] and confers to genes a transcriptional state responsive to developmental or environmental cues [15]. By reanalyzing publicly available datasets, we observed that 23% and 22% of the H3K27me3-marked TEs also harbor H2A.Z or the H2AKub mark, respectively (Fig. 2A). Interestingly, H3K27me3 long TEs are most often marked by H2A.Z and/ or H2Aub (Fig. 2A and B). Hence, they behave as PcG genic targets, which display the epigenetic hallmarks of silencing yet responsiveness (Fig. 2C).

Previous studies showed that, for a subset of genes, recruitment of H3K27me3 is dependent on the presence of H2A.Z [46] or H2AK121ub [50]. This prompted us to analyze H3K27me3 profiles at TEs in *arp6* mutant impaired for H2A.Z deposition, as well as in *atbmi* PRC1 mutants impaired in H2AUb deposition. At a subset of TEs, we observed a decrease of H3K27me3 in *arp6* (Fig. 2D, Additional file 1: Fig. S2A–B) and *atbmi* plants (Fig. 2E, Additional file 1: Fig.S2C–D), thus revealing similar dependencies for H3K27me3 deposition as for genes. Moreover, given that REF6 was shown to preferentially remove H3K27me3 at H2AK121ub-marked genes [49], we re-analyzed published data in the triple mutant *ref6 elf6 jmj13* (ELF6 and JMJ13 are known to function redundantly with REF6) [52]. We observed that several long TEs marked by H3K27me3 have higher levels of H3K27me3 in *ref6 elf6 jmj13* in comparison to

(See figure on next page.)

Fig. 2 H3K27me3 at TEs can be associated with H2AUb and H2A.Z, active demethylation, and transcriptional repression. **A**. Stacked bar plot showing distribution of TE marked by, either the H2A.Z variant or H2AK121 ubiquitination, among TE marked by H3K27me3 and per size range. A TE is considered as co-marked if both H3K27me3 and H2A.Z or H3K27me3 and H2AK121ub peaks overlap. **B**. Venn diagram showing the presence of H3K27me3, H2A.Z variant, and H2AK121ub on TEs > 3.5 kb. **C**. Representative genome browser view of 2 TEs from class 3 (marked by H3K27me3) with CHIP-seq data showing H2AK121ub (light green), H2A.Z (dark green), H3K27me3 (red) marks in wild-type (WT). **D**. Metagenes showing levels of H3K27me3 (top panel) and H2A.Z (bottom panel) in both WT and *arp6* mutant defective for H2A.Z incorporation at the TE subset impacted by the mutation. **E**. Metagenes showing levels of H3K27me3 (top panel) and H2AK121ub (bottom panel) in both WT and *atbmi1a/b/c* mutant defective for H2AK121ub deposition at the TE subset impacted by the mutation. **F**. Metagenes showing level of H3K27me3 in both WT and *ref6 elf6 jmj13* triple H3K27me3 demethylases mutant. The Left panel represents TEs already marked by H3K27me3 in WT but the mark is increased in *ref6 elf6 jmj13*. The right panel represents TEs that are not marked by H3K27me3 in WT but the mark is revealed in *ref6 elf6 jmj13*. **G**. Representative genome browser view of 2 TEs from class 3 showing H3K27me3 (red) marks in wild-type and *ref6 elf6 jmj13*. **H**. Representative browser view of a TE showing H3K27me3 (red) and RNA (blue) in WT and *swn clf* double mutant. **I**. Boxplot showing TEs from class 3 (3161 out of 4656) significantly up-regulated in *swn clf* compared to WT. Statistical analysis was performed using Wilcoxon-Mann–Whitney *U*-test

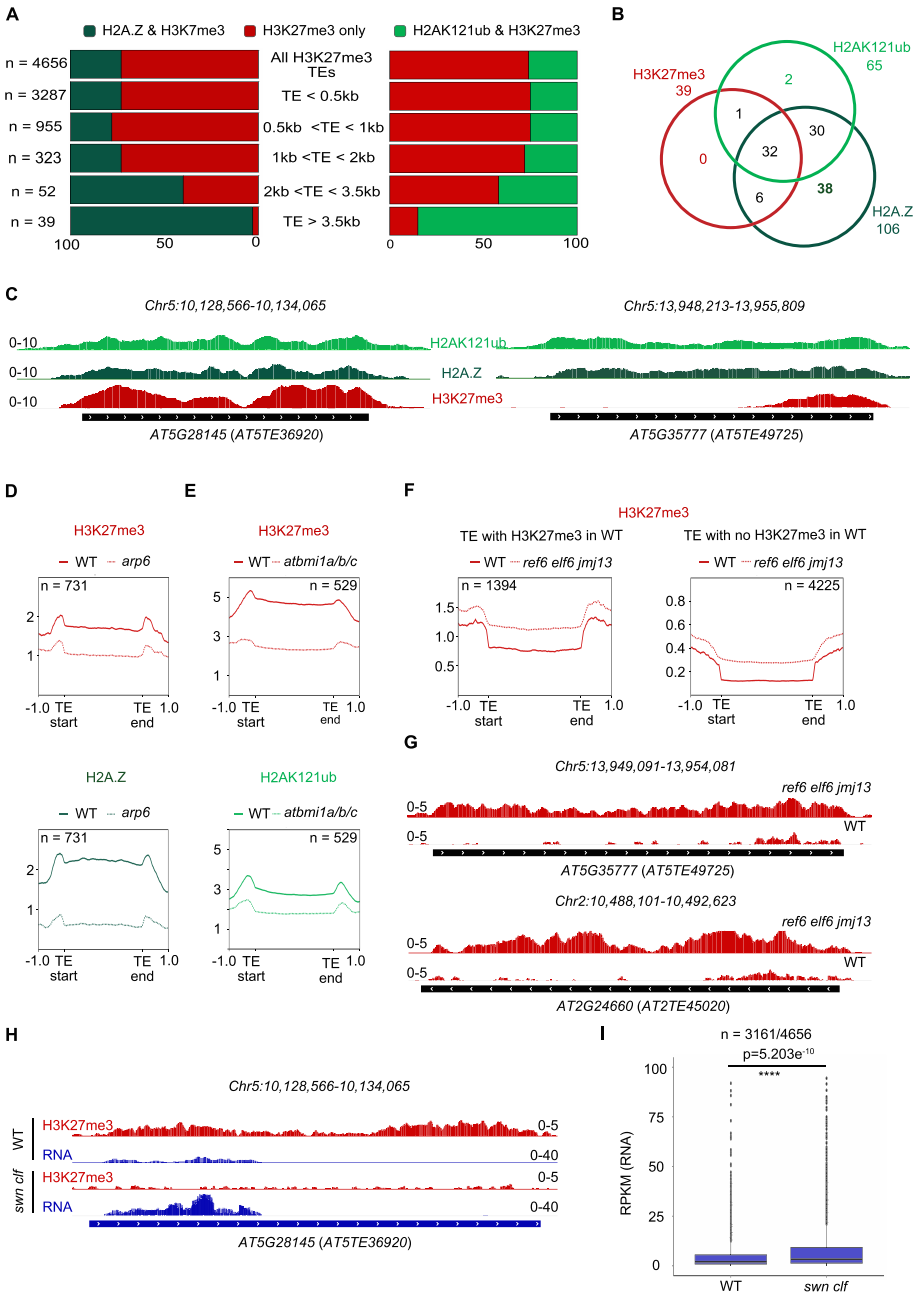

Fig. 2 (See legend on previous page.)

WT (Fig. 2F left panel, Fig. 2G and Additional file 1: Fig. S2E). Thus, the observed profile of H3K27me3 at some TEs is shaped and regulated through active demethylation by JMJ H3K27me3 demethylases (Fig. 2G). Interestingly, we also observed a small gain of H3K27me3 in the triple mutant at TEs marked by H2AK121ub but not by H3K27me3 in the WT (55% of cases) (Fig. 2F right panel, Additional file 1: Fig. S2F–G); this shows that the activity of H3K27me3 demethylases can remove H3K27me3 completely so that TEs look like active genes in WT.

These results demonstrate that TEs can harbor chromatin hallmarks of typical PcG genic targets, including active regulation by histone demethylase. These features distinguish them from constitutively heterochromatic TEs.

Finally, to assess the relevance of this H3K27me3 mark on TE, we exploited previously published RNA seq analyses comparing gene expression in WT versus *swn clf* PRC2 double mutant [53], in which virtually all H3K27me3 marks are erased in vegetative tissues (Fig. 2H, Additional file 1: Fig. S2H). We found that two-thirds of H3K27me3-marked TEs (3161 out of 4656) were significantly up-regulated in the double mutant compared to wild-type (Fig. 2I, Additional file 1: Fig. S2H). As for the remainder, lack of H3K27me3 may not be sufficient to activate the TEs, possibly because these TEs need transcriptional conditions such as transcription factor presence that are not met in *swn clf* (in line with the observation that loss of epigenetic silencing marks is not always sufficient to activate TE transcription [3, 54]). Thus, we conclude that H3K27me3 at TEs can transcriptionally repress their expression in Arabidopsis as shown in algae and bryophytes [37].

H3K27me3 mark is recruited at neo-inserted transgenic TEs in a *cis*-determined manner

Since TEs can recruit PcG in the absence of DNA methylation, we envisioned that PcG targeting could occur de novo at newly inserted TE copies, possibly in competition with DNA methylation. To test whether newly inserted TE copies devoid of DNA methylation upon insertion can recruit PcG, we used LTR retrotransposon-based transgenes which can model TE insertion [55]. We further reasoned that such transgenes could enable us to validate the hypothesis raised by Fig. 1, which is that recruitment of PcG on TE is instructed by intrinsic features of the TE. For this purpose, we used three *COPIA* LTR retrotransposon candidates that were chosen based on their genomic and epigenomic features. *ATCOPIA12B* (*AT2TE45020*) was identified in Fig. 1H as part of a family prevalently targeted by PcG in WT and harboring H3K27me3 (Additional file 1: Fig. S3A top panel), particularly in *ref6 elf6 jmj13* mutants in which the masking effect of REF6 activity is eliminated (Fig. 2G). *ATCOPIA21A* (*AT5TE65370*) is an active TE that is DNA methylated in WT but targeted by PcG in *ddm1* with a peak centered on the TE –suggesting that H3K27me3 does not come from nearby gene (Additional file 1: Fig. S3A bottom left panel) and was previously shown to transpose in *ddm1* mutant lines as well as in *ddm1*-derived EpiRILS [56, 57]. Finally, *ATCOPIA13A* (*AT2TE23885*) is DNA methylated in WT and targeted by PcG in *ddm1*, with different H3K27me3 patterns compared to *ATCOPIA21* and possibly with a contribution of H3K27me3 spreading from the nearby gene (Additional file 1: Fig. S3A bottom right panel).

The three sequences were cloned separately in binary vectors and a SNP was introduced to discriminate transgenes from endogenous sequences (Fig. 3A). We exploited the SNP to discriminate and quantify H3K27me3 levels on the transgene as compared to the endogenous original sequence. Besides, the H3K27me3 analysis was performed in pools of 15–20 plants to dilute positional effects linked to the transgene insertion site, in particular, H3K27me3 spreading from nearby regions. Results show consistent recruitment of H3K27me3 de novo on the transgenic TEs, although to different extents as quantified using reads counts (Fig. 3B–C, Additional file 1: Fig. S3B–D, right panels). While H3K27me3 spreading from regions adjacent to transgenic copies is possible and

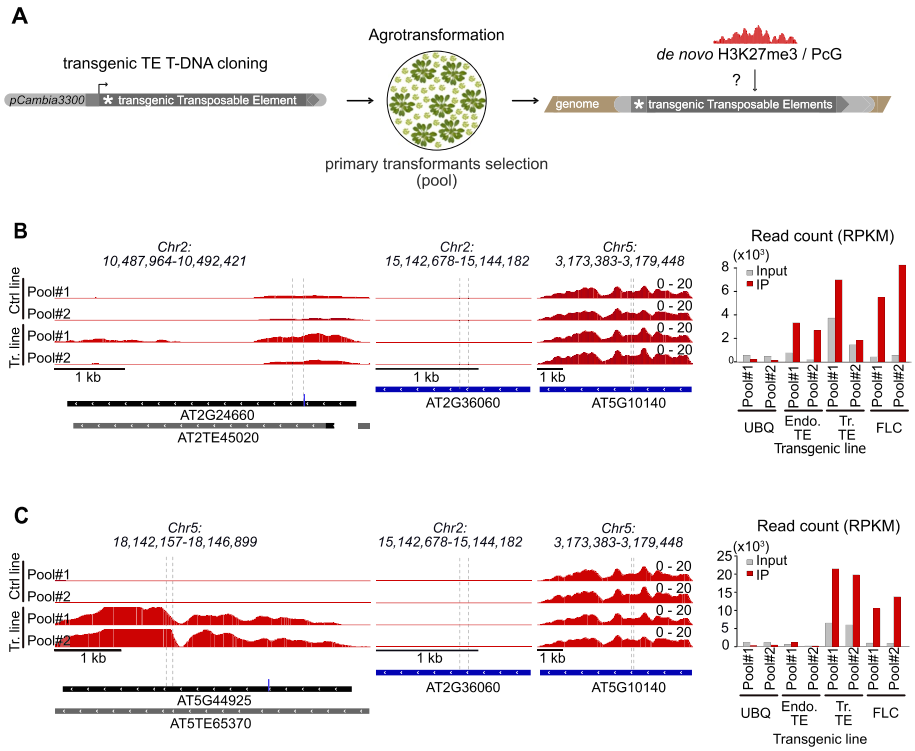

Fig. 3 Neo-inserted TE sequences can recruit H3K27me3 de novo and in an instructive manner. **A**. Experimental scheme: individual TE sequences were cloned into a binary vector, introduced in the plant genome, and studied by H3K27me3 ChIP-seq in large pools of T1 (15–20 plants) to dilute position effects. **B**. Left panel: representative genome browser view of H3K27me3 average profile (IP-INPUT) at *COPIA12B* sequence (top panel) in two independent pools of plants either transformed with unrelated transgene (Ctrl) or with the TE of interest. The H3K27me3 signal in the "Ctrl" tracks reflects the H3K27me3 status at the endogene TE. The H3K27me3 signal in the "Tr." tracks reflects the H3K27me3 status at the endogene + transgene: higher signal compared to the control line indicates recruitment of H3K27me3 on transgenic TEs in addition to the endogene. Variability between transgenic pools reflects the difference in transgenic TE copy number which varies between Pool 1 and 2. Middle panel: *UBQ* and *FLC* are shown are negative and positive controls respectively for H3K27me3 recruitment. Right panel: Recruitment on the transgene is determined by quantification of endogenous and transgenic *COPIA12B* sequences immunoprecipitated with H3K27me3 at the region where an SNP was introduced. ChIP-seq reads overlapping with the SNP at the *COPIA12B* locus were extracted, counted, and normalized by total read number. Recruitment on the transgene is validated by enrichment observed in IP as compared to an Input fraction. At control regions (*UBQ* and *FLC*), reads were extracted at positions *Chr5:3,178,750* (1st intron) and *Chr2:15,143,214* respectively (dashed lines). Copy number variation between transgenic pools can be visualized by the input DNA fraction (gray bar)/in the non-transgenic lines, variation of input reflects the difference in sequencing coverage at the SNP. **C**. Left, middle, and right panels are as in **B** for the *COPIA21* sequence

could contribute to the signal, the reproducible and consistent differences in H3K27me3 patterns between the three TEs indicate that the recruitment of H3K27me3 is specific to each TE (Fig. 3B–C, Additional file 1: Fig. S3B–D, left panels).

Switching between DNA methylation and Polycomb at TEs at the population level

The observation of relatively intact H3K27me3-marked TEs, which have gene-like chromatin features, raises the question of whether this epigenetic state at TE is conserved across accessions given that H3K27me3 genic targets are conserved across Arabidopsis species [16].

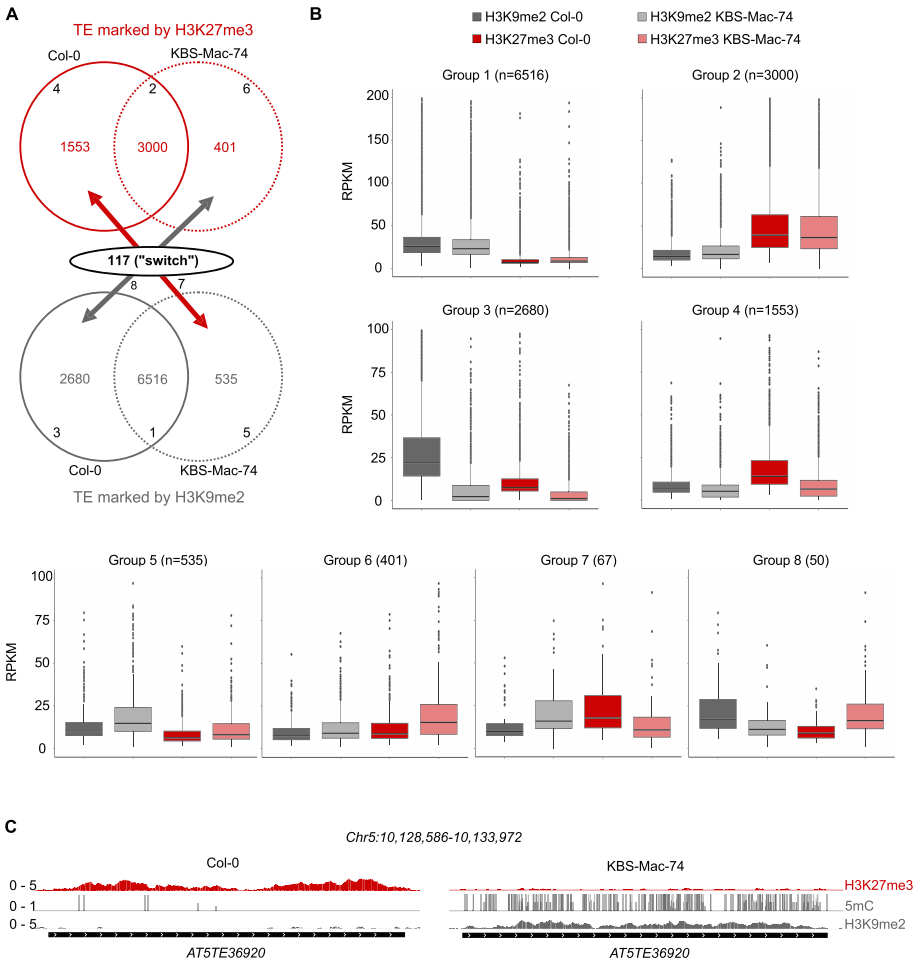

Fig. 4 Switch between silencing epigenetic marks at TEs visualized by inter-accessions comparisons. **A**. Venn diagram of TEs marked by H3K27me3, at the top (numbers in red), or by DNA methylation (H3K9me2 used as proxy) at the bottom (numbers in gray), in 2 different accessions, Col-0 and KBS-Mac-74. Intersection between both Venn diagrams shows TEs (107) that switch from H3K27me3 in one accession to H3K9me2 in the other. **B**. Box plots showing the levels of H3K9me2 in Col-0, KBS-Mac-74 and the levels of H3K27me3 in Col-0, KBS-Mac-74 (in RPKM) for the eight different TE-clusters identified in A. **C**. Representative genome browser view of H3K27me3 and DNA methylation levels at an orthologous TE copy in two accessions, Col-0 and KBS-Mac-74, respectively (red: H3K27me3, dark gray: CG methylation, gray: CHG methylation, light gray: CHH methylation, black bar: TE annotation)

To study this, we first compared Col-0 to KBS-Mac-74—an accession that was recently profiled for H3K27me3, H3K9me2, and DNA methylation [38] and for which high-quality ONT-based genome assembly was available [58]. Using these data, eight clusters of TEs were defined based on the length coverage of H3K27me3 and H3K9me2 profiles (Fig. 4A–B, Additional file 3). We chose to use H3K9me2 here as a proxy for DNA methylation for simplicity of representation and stringency of analysis: given that H3K9me2 is the hallmark of constitutive heterochromatic state mechanistically coupled with non-CG DNA methylation [5, 59] but also CG methylation [60] and that CG and CHG methylation are also coupled at heterochromatic loci [5], we reasoned that H3K9me2 variation (presence/absence) between accessions would be a relevant readout for a strong variation in epigenetic silencing state if any. In the two largest clusters (1 and 2) these two

marks remain unchanged; clusters 3 and 4 display epigenetic variations which are likely contributed by genetic diversity between both accessions (i.e., TEs that are present in Col-0 and absent in KBS-Ma-74, change in copy number and/or TE size); clusters 5 and 6 show variation of one mark but not the other. Strikingly, the last 2 clusters (7 and 8) (67 and 50 TEs, respectively) correspond to TEs with a change in epigenetic status between the two compared accessions (Fig. 4B). Thus, we have unveiled a sub-population of TEs with an epigenetic status that is variable at the species-level (Fig. 4C, Additional file 1: Fig. S4A for additional examples): we coin these TE as "*Bifrons*" which literally means "two-faced" in Latin. This phenomenon observed at orthologous TE insertions indicates a change in the epigenetic state—from standard TE-like DNA methylation to H3K27me3 or vice-versa, which we refer to as an "epigenetic switch". Interestingly, the epigenetic switch can be specific to the TE locus, with no observable epigenetic or transcriptional change at nearby regions (Additional file 1: Fig. S4B).

Other pairwise comparisons such as Col-0 versus Fly-2 or KBS-Mac-74 versus Fly-2, showed a similar number of TEs that differ in epigenetic state (Additional file 1: Fig. S4C). Based on this, we conclude that the *Bifrons* are a rather small sub-population of TEs, but they likely occur in all natural accessions of Arabidopsis.

H3K27me3 deposition at some TEs instead of DNA methylation is genetically linked to TE *cis*-determinants as well as *trans* factors

Next, we wanted to identify possible genetic determinants of this epigenetic variation. For this purpose, we exploited publicly available DNA methylation data for hundreds of Arabidopsis natural accessions and performed a Genome-Wide Association Study (GWAS). First, we took the opportunity that H3K27me3 and H3K9me2/DNA methylation data were recently profiled in 12 accessions [38] to test whether the presence of DNA methylation/H3K9me2 is generally correlated with the absence of H3K27me3 and vice-versa. To do this, for six individual accessions, we selected the TEs that have either a H3K9me2 or a H3K27me3 peak and observed that the large majority of TEs has either one mark or the other (Additional file 1: Fig. S5A). Accordingly, TEs with a H3K9me2 peak had low levels of H3K27me3 (Additional file 1: Fig. S5A, left metaplot) and TEs with a H3K27me3 peak had low levels of H3K9me2 (Additional file 1: Fig. S5A, right metaplot). This allows us to extrapolate a H3K27me3-marked status from the absence of DNA methylation (absence of H3K9me2) at TEs in nearly 500 accessions for which DNA methylation but not H3K27me3 data was available [61]. Thus, we could perform GWAS at a given TE, using the presence/absence of DNA methylation as a molecular phenotype.

We focused our analysis on two individual TEs of the *COPIA12* family that display genetic and epigenetic variation. *ATCOPIA12D* (*AT5TE36920*) insertion (shown in Fig. 4C) is present in all the assembled accessions while *ATCOPIA12B* (*AT2TE45020*) insertion is either present or absent and there can be multiple insertions of the corresponding sequence. Both TE sequences can be found either in a DNA methylated state and/or H3K27me3 state. DNA methylation level was estimated for CG, CHG, and CHH respectively in each accession for the two TEs (Fig. 5A). At *ATCOPIA12D* (*AT5TE36920*), a clearly bimodal distribution of DNA methylation was observed: most of the accessions are DNA methylated in all three contexts but a small fraction of

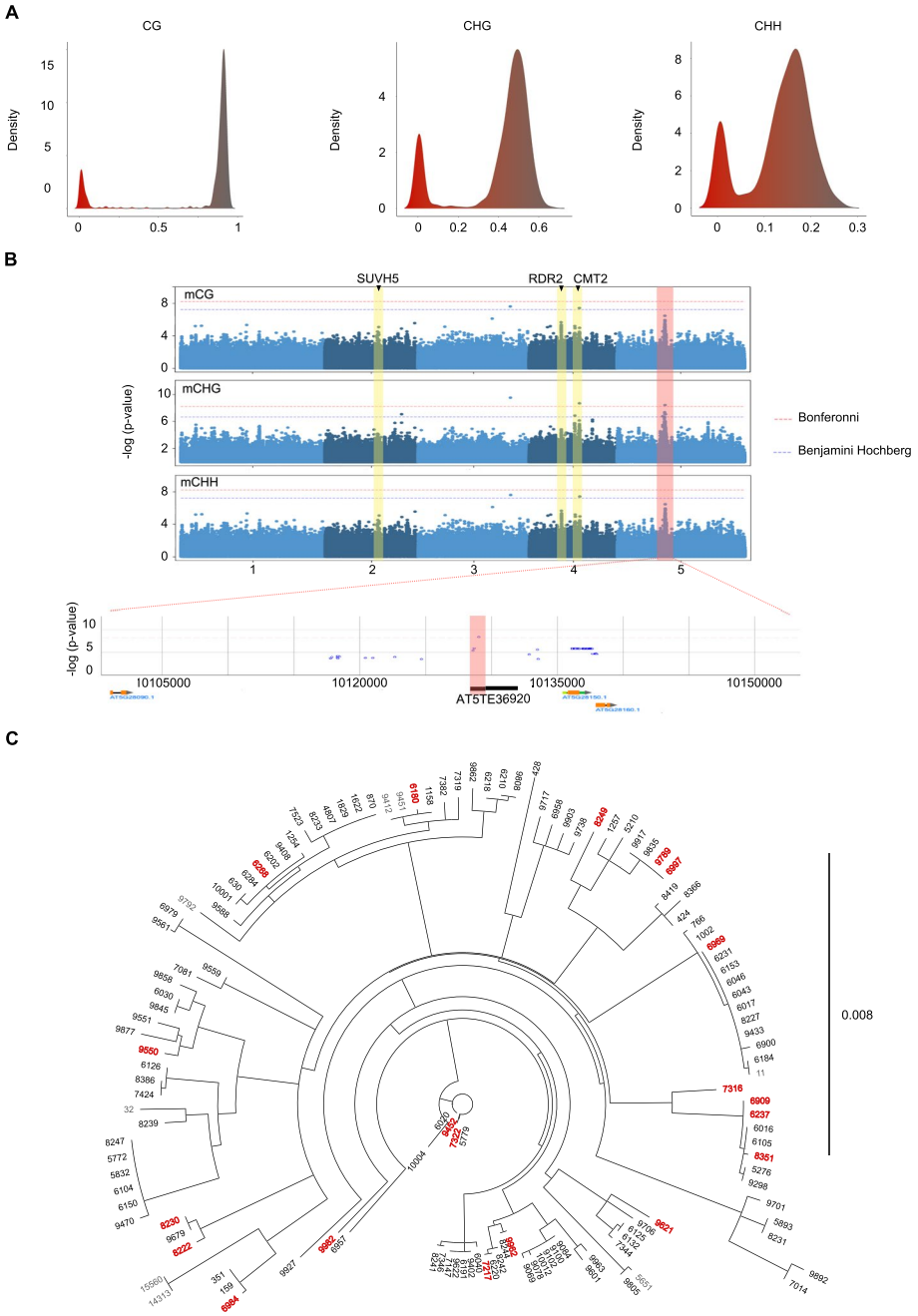

Fig. 5 Genetic determinants of the epigenetic switch. **A**. Plot showing DNA methylation value for an orthologous TE, *COPIA12D* (*AT5TE36920*), and the density across accessions in the 3 contexts, CG, CHG, and CHH. **B**. Manhattan plots show univariate GWAS mCG (upper panel), mCHG (middle panel), and mCHH (lower panel) for *COPIA12D* (*AT5TE36920*). Horizontal lines show genome-wide significance ($p = 0.05$ after Bonferroni correction or after Benjamini Hochberg). Yellow boxes indicate peaks in *trans*-factors related to DNA methylation and pink boxes indicate *cis*-region of the TE. **C**. Phylogenetic tree of *AT5TE36920* sequences in 130 different accessions. The genetic distance is indicated on the right. Accessions colored in red and bold: *COPIA12D* is devoid of DNA methylation and assumed to be marked by H3K27me3. Accessions colored in black: *COPIA12D* is DNA methylated. For accessions colored in gray, there is no information available about the epigenetic status of *COPIA12D* (based on [61])

accessions has methylation levels close to zero (Fig. 5A), likely with H3K27me3 instead as suggested by our previous results (Fig. 4). GWAS analysis showed a distinctive peak *in cis* located on the TE itself (Fig. 5B). In addition, other smaller peaks could be observed close to genes known to be important players of the DNA or H3K9 methylation pathway, such as *RDR2*, *CMT2*, and *SUVH5* (Additional file 1: Fig. S5B). Of note, CMT2 has been previously shown to be linked with DNA methylation natural variation at many TEs [62], and we found that these included *ATCOPIA12D*. At *AT2TE45020* (*ATCOPIA12B*), a bimodal distribution of DNA methylation was observed for the CG context in particular, most of the accessions being unmethylated (Additional file 1: Fig. S5C). GWAS for this TE again showed an association with a region encompassing the TE itself (Additional file 1: Fig. S5D). The observation of *cis*-peaks for both GWAS suggests that the epigenetic status could be *cis*-determined: this would support the instructive mode of H3K27me3 recruitment evidenced in Fig. 3. While we cannot exclude the possibility that H3K27me3 was not determined by a particular change in the TE sequence, but rather occurred by chance in some accessions and then was maintained by robust epigenetic inheritance common in plants [63], our insertion experiments described above that included *ATCOPIA12B* support the link between the TE sequence and the establishment of H3K27me3. Moreover, the identification of *trans*-factors further indicates that the switch is likely multifactorial.

To get more insight into the nature of possible *cis*-determinants, we extracted the sequence of *COPIA12D* from the fully assembled genomes of 130 accessions extracted (kindly provided by the Nordborg lab, GMI) (Additional file 4). We aligned them to build a phylogenic tree and observed that H3K27me3-marked sequences did not cluster together (Fig. 5C). This shows that there are no common haplotypes causal to the epigenetic switch to H3K27me3, and further indicates that the switch has occurred several times. Despite the presence of shared SNPs as revealed by GWAS, we could not find by de novo search any sequence motif sequences that were significantly enriched in H3K27me3 and that would be absent in the DNA methylated ones by motif search analysis. Likewise, none of the PRE (Polycomb responsive elements)-like motifs previously published to be important for PcG recruitment at genes [64–68] was enriched in H3K27me3-marked sequences (Additional file 1: Fig. S5E) neither were CG, CHG, or CHH di/tri-nucleotides (Additional file 1: Fig. S5F). One likely explanation for this is that some DNA methylated sequences could have the specific cis-elements needed to recruit H3K27me3 but this potential to recruit H3K27me3 is masked by the presence of DNA methylation; this hypothesis could be tested in the future by profiling H3K27me3 marks in DNA methylation mutants such as *met1* or *ddm1* in Arabidopsis accessions [69]. For now, this scenario is supported by our previous observations of enrichment for cis-elements in heterochromatic TEs that gain H3K27me3 in ddm1 versus all heterochromatic TEs [23]. At least, we can conclude here that there are no detectable, discrete cis-motifs that are necessary and sufficient to make a TE target of PcG instead of DNA methylation. This is in agreement with GWAS showing that the determination of TE epigenetic status is multifactorial and involves trans-factors involved in DNA methylation.

Thus, the epigenetic switches are the product of multifactorial causes leading to a change in the TE silencing mode and have likely occurred independently in nature during the species evolution.

Discussion

In higher plants, TEs marked by H3K27me3 in wild type have recently gained interest. While many H3K27me3-marked TE relics have been recently described [37], here we report the existence of TEs that can be intact and are targeted by Polycomb instead of DNA methylation. This is important because it suggests a biological role of PcG pathway as an alternative silencing system which could prevent transposition and/or regulate the production of TE-derived proteins with a function for the host. The functionality of H3K27me3 as an alternative silencing mark at these intact TEs is indeed supported by the increase of transcription that we observed upon loss of H3K27me3 in *swn clf* mutants. Furthermore, our study provides multiple lines of evidence that the presence of H3K27me3 at these TEs may be linked to *cis*-determinants as shown for genes [67]. First, we observed that long H3K27me3-TEs can be found in all genomic compartments, including pericentromeric regions. Second, a transgenic approach showed that three different TE sequences display different and specific patterns of H3K27me3 upon neo-insertion. This demonstrates that PcG recruitment can be instructed by the TE copies themselves independently of position effects. Finally, and in support of this, we found that an orthologous TE copy can switch epigenetic status across accessions, independently of any epigenetic and transcriptional changes in the neighboring regions; GWAS analyses further support a link between the TE sequence and the establishment of H3K27me3. The intrinsic ability of long TE copies to recruit PRC2 raises the possibility that during the process of TE erosion, H3K27me3 nucleation regions could have been selected. Interestingly, short H3K27me3-TE relics were preferentially found nearby genes [37]. Future work should investigate whether these relics could act as epigenetic modules that can impact gene expression, as shown for DNA-methylated TE remnants.

The observation that many TEs are marked by H3K27me3 raises the question of why TEs would be targeted, at times, by a plastic silencing mechanism instead of DNA methylation. When they first insert into the genome, TEs (in particular retro-elements) are DNA hypomethylated since neo-synthesized DNA (retrotranscribed) is inserted: the targeting of PcG at this stage could allow for rapid silencing of the element while DNA methylation is being established progressively. A biological role of PcG silencing is also supported by co-occurrence with H2A.Z and H2Aub and the active modulation of H3K27me3 patterns by REF6/ELF6/JMJ13. As shown for H3K27me3-marked genes, this may contribute to a partially repressed state of the TE that enables a quick response to environmental or developmental cues. This may be important for organisms at certain windows of their life cycle as exemplified by TEs [70, 71] which expression is necessary during embryonic development in mammals.

Finally, inter-accession comparisons uncovered repeated epigenetic switching between H3K27me3 and DNA methylation or vice-versa, at a small population of TEs. Future investigations should reveal in which context such switches are triggered and in which direction, although we predict that recapitulating the switch experimentally is going to be challenging. Yet, several non-mutually exclusive scenarios can already be proposed (Fig. 6A–C). One scenario ("*switch 1*": from H3K27me3 to DNA methylation) is that certain TEs have features allowing them to be targeted by Polycomb upon insertion in the genome; this is supported by our results with transgenic

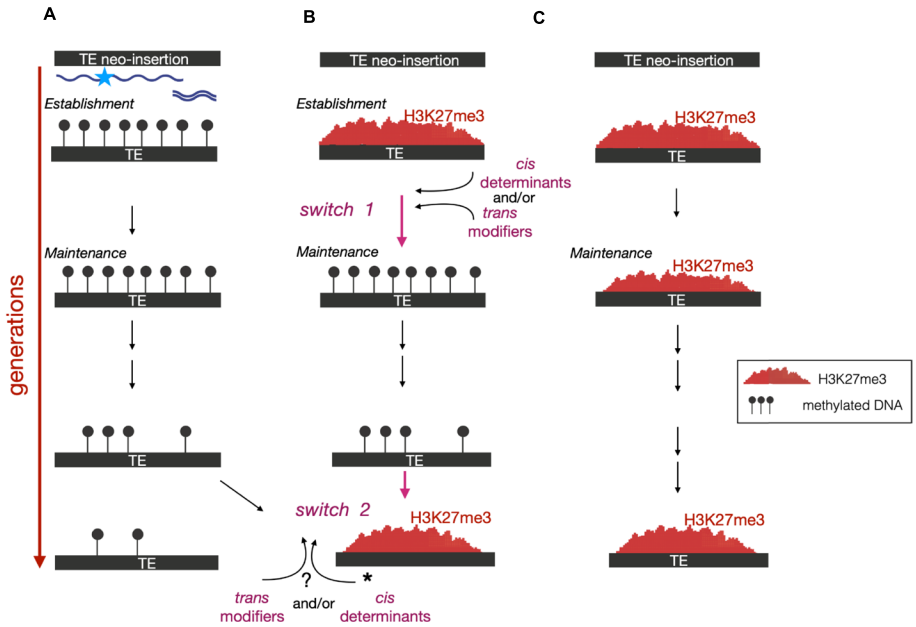

Fig. 6 Scenario for epigenetic switch throughout TE lifetime. **A**. So far, previous studies have established a model whereby TEs arrive in the genome after a burst of transposition or horizontal transfer, and are detected by the RNAi machinery following their transcription or by pre-existing homologous small RNAs. DNA methylation is established by small RNA-directed DNA methylation (RdDM) and maintained by a combination of RdDM and RNA-independent mechanisms. Eventually, TE decays and can keep passively DNA methylation or lose it. The results obtained in this study lead to propose additional phases to this scenario of silencing progression, at least for some elements, which could comprise one or two switches "DNA methylation-H3K27me3". **B**. One scenario could be that PcG would be recruited first, either alone or through DRM2. This targeting could slow down the onset of DNA methylation yet DNA methylation would come in at some point, amplify through generations, and become dense enough to antagonize PRC2—this would constitute a first possible "switch 1"—and this silencing state more stable hence found more frequently. There could be another, not mutually exclusive, switch ("switch 2") when, as part of his decaying process/by mutating, the TE loses the capacity to maintain DNA methylation and /or mutates to promote PcG recruitment. Eventually, some of these H3K27me3-TEs could also be selected for structural or regulatory functions. **C**. Some TEs may gain H3K27me3 upon insertion in the genome and keep this epigenetic status throughout their lifetime

TEs able to recruit H3K27me3 de novo. Then at some point, upon TE detection by the RNAi machinery leading to the onset of RdDM, and/or DNA methylation amplification, PRC2 would be antagonized ("*switch 1*", Fig. 6B) to reach a "DNA-methylated only" epigenetic status. One other possible switch ("*switch 2*": from DNA methylation to H3K27me3) would entail prior loss of DNA methylation through loss of *cis* methylated cytosines and/or expression changes of *trans* modifiers, combined with *cis* molecular determinants of PcG recruitment such as PREs ("*switch 2*", Fig. 6A–B). Meeting all these requirements could take many generations. Of note, in our analyses, only a fraction of the host TEs have undergone the switch, which likely indicates that they are in a transient evolutionary state. It will be important in the future to understand how TE silencing progresses from one system to the other, because DNA methylation and H3K27me3 do not have the same properties in terms of maintenance and silencing stability. Hence, this trajectory would impact the TE potential for mobilization and domestication as regulatory modules, which could profoundly impact the integrity and expression of host genomes.

Conclusions

In conclusion, the data presented in this paper challenge existing notions about the relationship between the two major modes of epigenetic silencing in eukaryotes by revealing (i) an alternative mode of TE silencing associated with H3K27me3 instead of DNA methylation in flowering plants which can be deployed at the various stages of TE lifespan, (ii) an epigenetic switching between them which can occur through evolutionary time. This epigenetic plasticity of TE control is a new paradigm in TE regulation, and more generally in epigenetics, which may well extend to other higher eukaryotes.

Methods

Generation of transgenic lines

TE sequences were synthesized and cloned in pUC57 by Genescript. TEs were subsequently cloned in pCAMBIA3300 using restriction enzymes and a T4 DNA ligase-based cloning strategy. Plants were transformed via *Agrobacterium tumefaciens* floral dip [72] and T1 plants selected on appropriate herbicide for selection.

Plant material and growth conditions

All experiments were conducted on A. *thaliana* grown in vitro for two weeks on ½ MS plates under short light-day conditions (8-h light/16-h dark photoperiod and 22 °C). The transgenic plants were selected on Basta after 2 weeks, transferred in soil and 4 weeks old rosette leaves were collected in pools of 15–20 plants to perform ChIP-seq.

Chromatin immunoprecipitation (ChIP)

ChIP experiments were conducted with 1 g of seedlings (Fig. 1) (two biological replicates) or 1 g of 4-week-old rosette leaves (Fig. 3) (two biological replicates for each transgenic TE studied). Briefly, plants were in vivo crosslinked with 1% formaldehyde solution supplemented with 1 mM PMSF to avoid protein degradation. Crosslinked chromatin was sonicated for 25 cycles (30 s ON/30 s OFF) at high intensity with Diagenode Bioruptor. ChIP was performed overnight using Invitrogen Protein A Dynabeads and anti-H3K27me3 antibody. Precipitated DNA was eluted and purified using phenol:chloroform. Untreated and sonicated chromatin was processed in parallel and considered as the input. DNA was sequenced (sequencing single-reads, 1×50 bp or paired-end 100 bp; Illumina) of the resulting input and IP samples performed by BGI.

Chromatin immunoprecipitation (ChIP)-seq data analysis

ChIP-seq data were analyzed using Nextflow (v20.04.1) and nf-core/chipseq pipeline (v1.2.2). Raw sequencing data were quality-checked using FastQC software (v0.11.7) and low-quality reads were removed Trim Galore (v0.6.5). Reads were mapped using BWA [73] onto TAIR10 A. *thaliana* and multi-mapped reads were kept. Genomic regions significantly marked by H3K27me3 were identified using MACS2 [74] and genes or TEs overlapping these regions were obtained using Bedtools [75]. ComputeMatrix from Deeptools was used to create a score matrix and to generate heatmeaps (plotHeatmap) and metaplots (plotProfile) as graphical output. To define lists of TEs that are differentially regulated between various mutants (to then analyze the changes of H3K27me3 levels in Fig. 2) or between accessions (to then be able to define different clusters in Fig. 4),

we used the presence or absence of peaks (together with length coverage) as generated by MACS2; to generate box-plots, we extracted the H3K27me3-ChIP-seq reads from the different TEs.

Read count analyses

Reads overlapping with the SNP (between transgene and endogene) at *ATCOPIA12B*, *ATCOPIA13A*, and *ATCOPIA21A* locus were extracted, counted, and normalized by total read number using Samtools view. At control regions (UBQ and FLC), reads were extracted at positions *Chr2:15,143,214* (UBQ) and *Chr5:3,178,750* (1st intron FLC), respectively.

Bisulfite-sequencing analyses

Raw sequencing data were quality-checked using FastQC software (v0.11.7). Adapter and low-quality sequences were removed from the raw data using Trim Galore (v0.6.5). The reference genome (TAIR10) was prepared for indexing using bismark_genome_preparation function of the Bismark software (v0.22.2) [76]. Mapping was performed using the Bowtie2 algorithm (v2.4.4) [77]. Mapped reads were deduplicated to remove PCR-induced duplicate reads using the deduplicate_bismark function of Bismark software. Methylation levels for each cytosine context (CG, CHG, and CHH) were extracted and analyzed from the mapped data using the bismark_methylation_extractor function of Bismark software. The resulting files were converted to Bedgraph format using the bismark2bedGraph function of Bismark software to conduct further analysis with Bedtools.

GWAS analyses

GWASs were performed online on: https://gwas.gmi.oeaw.ac.at with default parameters and without transformation.

Analyzed data sets

The public datasets were downloaded from NCBI GEO using the specified GEO accession numbers: H2A.W and H2A.Z ChIP-seq (NCBI GEO GSE50942) [78], BS-seq Col-0 (NCBI GEO GSE226560) [38], WT H2AK121ub and *atbmi1abc* H3K27me3 and H2AK121ub ChIP-seq (NCBI GEO GSE89358) [50], *arp6* H3K27me3 and H2AZ (NCBI GEO GSE103361) [46], *ref6 elf6 jmj13* H3K27me3 ChIP-seq (GSE106942) [52], WT and *swn clf* H3K27me3 ChIP-seq and RNA-seq (NCBI GEO GSE108960) [53], accessions related data (NCBI GEO GSE224761) [38]. The datasets produced as part of this study are accessible on ENA (European Nucleotide Archive) (PRJEB83093, ERP166759) [79].

Phylogenetic tree and multiple alignments

For *COPIA12D*, sequences were extracted from 130 ONT-based genome assemblies of Arabidopsis accessions (kindly provided by the Nordborg lab, GMI) and are available in Additional file 4. The sequences were aligned using the MUSCLE default parameter in Unipro UGENE. Then the phylogenetic tree derived from the sequences was made by the PHYLYP Neighbor Joining Method in Unipro UGENE.

Transposable element domain analysis

The domains of LTR Copia transposable elements were predicted using Domain-based Annotation of Tranposable Elements (DANTE) [39] available on RepeatExplorer web server http://repeatexplorer.org/.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-024-03466-6.

Additional file 1: Fig. S1-S5.

Additional file 2: Table S1. List of TEs of each class presented in Fig. 1.

Additional file 3: Table S2. List of TEs in each cluster presented in Fig. 4.

Additional file 4: Table S3. List of the 130 COPIA12D sequences extracted from assembled natural accessions.

Additional file 5. Peer review history.

Acknowledgements

This work was funded by the French State Grant ANRJCJC (ANR-19-CE12-0033-01 to A.D) and has benefited from a grant Saclay-Plant Sciences (SPS) (ANR-17-EUR-0007, EUR SPS-GSR) under a France 2030 program (ANR-11-IDEX-0003), as well as from support from the Genome Biology department of Institut de Biologie Intégrative de la Cellule (I2BC). We thank Bouché N. and Toffano-Nioche C. for their help with bioinformatic analyses as part of the SPS bioinformatic work group. We thank the services and platforms of I2BC for excellent technical support (in particular A. Almeida and F. Culot for their assistance) and V. Couvreux for excellent plant care. We also thank our colleagues for discussions, in particular the CNRS Groupements De Recherche "Epiplant" (GDR2027) and "Mobil'ET (GDR3546).

Review history

The review history is available as Additional file 5.

Peer review information

Robert Kofler and Tim Sands were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Authors' contributions

AD conceived the study. VH and AD wrote the paper. FPP and TY performed experiments and analyses related to transgenic TE sequences; VH performed the remainder of the experiments and bioinformatic analyses. CV contributed to bioinformatics analyses, in particular phylogenetic analyses of TEs. GA contributed to customized bioinformatic analysis of transgenic plants. MN shared H3K27me3/H3K9me2/DNA methylation profiles and ONT assemblies in natural accessions ahead of publication. AEK contributed to GWAS analyses and sequence extraction for phylogenetic analyses. All authors contributed to various extents to data interpretation and writing of the manuscript.

Data availability

Access to analyzed datasets is provided in the methods; the datasets produced as part of this study are available in the European Nucleotide Archive (ENA) repository under accession number PRJEB83093[78]. The corresponding author should be contacted for the request of other data and material related to this study.

Declarations

Ethics approval and consent to participate

Not applicable to this study.

Competing interests

The authors declare that they have no competing interests.

Received: 15 April 2024 Accepted: 19 December 2024
Published online: 20 January 2025

References

1. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science. 2016;351:1083–7.
2. Liu P, Cuerda-Gil D, Shahid S, Slotkin RK. The Epigenetic Control of the Transposable Element Life Cycle in Plant Genomes and Beyond. Annu Rev Genet. 2022;56:63–87.
3. Zervudacki J, Agnès Y, Delase A, Navarro L, Deleris A. Transcriptional control and exploitation of an immune-responsive family of plant retrotransposons. EMBO J. 2018;37:e98482.
4. Johnson LM, Cao X, Jacobsen SE. Interplay between Two Epigenetic Marks: DNA Methylation and Histone H3 Lysine 9 Methylation. Curr Biol. 2002;12:1360–7.

5. Du J, Johnson LM, Jacobsen SE, Patel DJ. DNA methylation pathways and their crosstalk with histone methylation. Nat Rev Mol Cell Biol. 2015;16:519–32.
6. Cao X, Jacobsen SE. Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci U S A. 2002;99:16491–8.
7. Cao X, Jacobsen SE. Role of the Arabidopsis DRM Methyltransferases in De Novo DNA Methylation and Gene Silencing. Curr Biol. 2002;12:1138–44.
8. Chan SW-L, et al. RNA Silencing Genes Control de Novo DNA Methylation. Science. 2004;303:1336–1336.
9. Kankel MW, et al. Arabidopsis MET1 cytosine methyltransferase mutants. Genetics. 2003;163:1109–22.
10. Lindroth AM, et al. Requirement of CHROMOMETHYLASE3 for Maintenance of CpXpG Methylation. Science. 2001;292:2077–80.
11. Stroud H, et al. The roles of non-CG methylation in Arabidopsis. Nat Struct Mol Biol. 2014;21:64–72.
12. Zhu J-K. Active DNA Demethylation Mediated by DNA Glycosylases. Annu Rev Genet. 2009;43:143–66.
13. Penterman J, et al. DNA demethylation in the Arabidopsis genome. Proc Natl Acad Sci U S A. 2007;104:6752–7.
14. del Prete S, Mikulski P, Schubert D, Gaudin V. One, Two, Three: Polycomb Proteins Hit All Dimensions of Gene Regulation. Genes. 2015;6:520–42.
15. Baile F, Gómez-Zambrano Á, Calonje M. Roles of Polycomb complexes in regulating gene expression and chromatin structure in plants. Plant Commun. 2021;3: 100267.
16. Chica C, Louis A, Roest Crollius H, Colot V, Roudier F. Comparative epigenomics in the Brassicaceae reveals two evolutionarily conserved modes of PRC2-mediated gene regulation. Genome Biol. 2017;18:207.
17. Marasca F, Bodega B, Orlando V. How polycomb-mediated cell memory deals with a changing environment: variations in PcG complexes and proteins assortment convey plasticity to epigenetic regulation as a response to environment. BioEssays News Rev Mol Cell Dev Biol. 2018;40:e1700137.
18. Vijayanathan M, Trejo-Arellano MG, Mozgová I. Polycomb Repressive Complex 2 in Eukaryotes—An Evolutionary Perspective. Epigenomes. 2022;6:3.
19. Lu F, Cui X, Zhang S, Jenuwein T, Cao X. Arabidopsis REF6 is a histone H3 lysine 27 demethylase. Nat Genet. 2011;43:715–9.
20. Crevillén P, et al. Epigenetic reprogramming that prevents transgenerational inheritance of the vernalized state. Nature. 2014;515:587–90.
21. Borg M, et al. Targeted reprogramming of H3K27me3 resets epigenetic memory in plant paternal chromatin. Nat Cell Biol. 2020;22:621–9.
22. Antunez-Sanchez J, et al. A new role for histone demethylases in the maintenance of plant genome integrity. eLife. 2020;9:e58533.
23. Rougée M, et al. Polycomb mutant partially suppresses DNA hypomethylation–associated phenotypes in Arabidopsis. Life Sci Alliance. 2021;4:e202000848.
24. Mathieu O, Probst AV, Paszkowski J. Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. EMBO J. 2005;24:2783–91.
25. Weinhofer I, Hehenberger E, Roszak P, Hennig L, Köhler C. H3K27me3 profiling of the endosperm implies exclusion of polycomb group protein targeting by DNA methylation. PLoS Genet. 2010;6:1–14.
26. Deleris A, et al. Loss of the DNA Methyltransferase MET1 Induces H3K9 Hypermethylation at PcG Target Genes and Redistribution of H3K27 Trimethylation to Transposons in Arabidopsis thaliana. PLoS Genet. 2012;8: e1003062.
27. Reddington JP, et al. Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. Genome Biol. 2013;14:R25.
28. Saksouk N, et al. Redundant Mechanisms to Form Silent Chromatin at Pericentromeric Regions Rely on BEND3 and DNA Methylation. Mol Cell. 2014;56:580–94.
29. Basenko EY, et al. Genome-wide redistribution of H3K27me3 is linked to genotoxic stress and defective growth. Proc Natl Acad Sci. 2015;112:E6339–48.
30. Walter M, Teissandier A, Pérez-Palacios R, Bourc'His D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. Life. 2016;5:1–30.
31. Guo W, Wang D, Lisch D. RNA-directed DNA methylation prevents rapid and heritable reversal of transposon silencing under heat stress in Zea mays. PLoS Genet. 2021;17: e1009326.
32. Brinkman AB, et al. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Res. 2012;22:1128–38.
33. Statham AL, et al. Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. Genome Res. 2012;22:1120–7.
34. Sato H, Santos-González J, Köhler C. Combinations of maternal-specific repressive epigenetic marks in the endosperm control seed dormancy. eLife. 2021;10:e64593.
35. Moreno-Romero J, Del Toro-De León G, Yadav VK, Santos-González J, Köhler C. Epigenetic signatures associated with imprinted paternally expressed genes in the Arabidopsis endosperm. Genome Biol. 2019;20:41.
36. Déléris A, Berger F, Duharcourt S. Role of Polycomb in the control of transposable elements. Trends Genet TIG. 2021;37:882–9.
37. Hisanaga T, et al. The Polycomb repressive complex 2 deposits H3K27me3 and represses transposable elements in a broad range of eukaryotes. Curr Biol CB. 2023;33:4367–4380.e9.
38. Kornienko AE, Nizhynska V, Molla Morales A, Pisupati R, Nordborg M. Population-level annotation of lncRNAs in Arabidopsis reveals extensive expression variation associated with transposable element-like silencing. Plant Cell. 2023;36:85–111.
39. Neumann P, et al. Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. Mob DNA. 2019;10:1.
40. Batista RA, et al. The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons. eLife. 2019;8:e50541.

41. Zhao Y, et al. Transposable elements: distribution, polymorphism, and climate adaptation in populus. Front Plant Sci. 2022;13:814718.
42. Barro-Trastoy D, Köhler C. Helitrons: genomic parasites that generate developmental novelties. Trends Genet. 2024;40:437–48.
43. Baile F, Calonje M. Dynamics of polycomb group marks in Arabidopsis. Curr Opin Plant Biol. 2024;80: 102553.
44. Jamge B, et al. Histone variants shape chromatin states in Arabidopsis. Life. 2023;12:RP87714.
45. Dai X, et al. H2A.Z represses gene expression by modulating promoter nucleosome structure and enhancer histone modifications in arabidopsis. Mol Plant. 2017;10:1274–92.
46. Carter B, et al. The Chromatin Remodelers PKL and PIE1 Act in an Epigenetic Pathway That Determines H3K27me3 Homeostasis in Arabidopsis. Plant Cell. 2018;30:1337–52.
47. Wang Y, et al. Histone variants H2A.Z and H3.3 coordinately regulate PRC2-dependent H3K27me3 deposition and gene expression regulation in mES cells. BMC Biol. 2018;16:107.
48. Gómez-Zambrano Á, Merini W, Calonje M. The repressive role of Arabidopsis H2AZ in transcriptional regulation depends on AtBMI1 activity. Nat Commun. 2019;10:2828.
49. Kralemann LEM, et al. Removal of H2Aub1 by ubiquitin-specific proteases 12 and 13 is required for stable Polycomb-mediated gene repression in Arabidopsis. Genome Biol. 2020;21:144.
50. Zhou Y, Romero-Campero FJ, Gómez-Zambrano Á, Turck F, Calonje M. H2A monoubiquitination in Arabidopsis thaliana is generally independent of LHP1 and PRC2 activity. Genome Biol. 2017;18:69.
51. Huang Y, et al. Evolution and conservation of polycomb repressive complex 1 core components and putative associated factors in the green lineage. BMC Genomics. 2019;20:533.
52. Yan W, et al. Dynamic and spatial restriction of Polycomb activity by plant histone demethylases. Nat Plants. 2018;4:681–9.
53. Shu J, et al. Genome-wide occupancy of histone H3K27 methyltransferases CURLY LEAF and SWINGER in Arabidopsis seedlings. Plant Direct. 2019;3: e00100.
54. Ito H, et al. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature. 2011;472:115–9.
55. Fultz D, Slotkin RK. Exogenous Transposable Elements Circumvent Identity-Based Silencing, Permitting the Dissection of Expression-Dependent Silencing. Plant Cell. 2017;29:360–76.
56. Tsukahara S, et al. Bursts of retrotransposition reproduced in Arabidopsis. Nature. 2009;461:423–6.
57. Quadrana L, et al. Transposition favors the generation of large effect mutations that may facilitate rapid adaptation. Nat Commun. 2019;10:3421.
58. Michael TP, et al. High contiguity Arabidopsis thaliana genome assembly with a single nanopore flow cell. Nat Commun. 2018;9:541.
59. Law JA, et al. Polymerase-IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature. 2013;498:385–9.
60. Rajakumara E, et al. A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. Genes Dev. 2011;25:137–52.
61. Kawakatsu T, et al. Epigenomic diversity in a global collection of Arabidopsis thaliana accessions. Cell. 2016;166:492–505.
62. Sasaki E, Gunis J, Reichardt-Gomez I, Nizhynska V, Nordborg M. Conditional GWAS of non-CG transposon methylation in Arabidopsis thaliana reveals major polymorphisms in five genes. PLoS Genet. 2022;18: e1010345.
63. Paszkowski J, Grossniklaus U. Selected aspects of transgenerational epigenetic inheritance and resetting in plants. Curr Opin Plant Biol. 2011;14:195–203.
64. Berger N, Dubreucq B, Roudier F, Dubos C, Lepiniec L. Transcriptional Regulation of Arabidopsis LEAFY COTYLEDON2 Involves RLE, a cis-Element That Regulates Trimethylation of Histone H3 at Lysine-27. Plant Cell. 2011;23:4065–78.
65. Qüesta JI, Song J, Geraldo N, An H, Dean C. Arabidopsis transcriptional repressor VAL1 triggers Polycomb silencing at FLC during vernalization. Science. 2016;353:485–8.
66. Yuan W, et al. A cis cold memory element and a trans epigenome reader mediate Polycomb silencing of FLC by vernalization in Arabidopsis. Nat Genet. 2016;48:1527–34.
67. Xiao J, et al. Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis. Nat Genet. 2017;49:1546–52.
68. Zhou Y, et al. Telobox motifs recruit CLF/SWN–PRC2 for H3K27me3 deposition via TRB factors in Arabidopsis. Nat Genet. 2018;50:638–44.
69. Srikant T, et al. Canalization of genome-wide transcriptional activity in Arabidopsis thaliana accessions by MET1-dependent CG methylation. Genome Biol. 2022;23:263.
70. Jachowicz JW, et al. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. Nat Genet. 2017;49:1502–10.
71. Yu H, et al. Dynamic reprogramming of H3K9me3 at hominoid-specific retrotransposons during human preimplantation development. Cell Stem Cell. 2022;29:1031–1050.e12.
72. Clough SJ, Bent AF. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J Cell Mol Biol. 1998;16:735–43.
73. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma Oxf Engl. 2009;25:1754–60.
74. Zhang Y, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9:R137.
75. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinforma Oxf Engl. 2010;26:841–2.
76. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinforma Oxf Engl. 2011;27:1571–2.

77. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25.
78. Yelagandula R, et al. The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in Arabidopsis. Cell. 2014;158:98–109.
79. Hure V, Piron-Prunier F, Yehouessi T. et al. Alternative silencing states of Transposable Elements in Arabidopsis associated with H3K27me3. ChIP-seq data on European Nucleotide Archive. 2024. https://www.ebi.ac.uk/ena/browser/view/PRJEB83093. Access Dec 6.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.