Review history

First round of review

Reviewer 1

While most transposable elements (TEs) are marked by DNA methylation, recent evidence revealed that the PRC2-mediated H3K27me3 modification can also target TEs in specific tissues or even preferentially in specific species. In this manuscript, the authors report the presence of H3K27me3-marked TEs in Arabidopsis and show that newly inserted TEs can be specifically de novo targeted by PRC2. They further show that the same TE can have alternative silencing states in different Arabidopsis accessions (DNA methylation or H3K27me3) and that the recruitment of H3K27me3 likely depends on specific cis elements. I found this manuscript interesting and could mostly follow the conclusions of the authors. Nevertheless, there are a few points that require clarification or additional analysis to solidify the conclusions.
Major points:
1. Line 141: The functional domains should be specified and included as a data table.
2. Figure 1B: The authors broadly show DNA methylation, but it would be relevant to specify the type of DNA methylation that overlaps with H3K27me3. This could possibly hint at how the switching or the establishment of H3K27me3 occurs.
3. Figure 1F: The authors find that mainly DNA TEs and helitrons are marked by H3K27me3. Since both types of TEs are frequently found to contain transcription factor binding sites and are part of promoters, one possible explanation for this observation could be that these TEs are close to genes and function as promoter elements. It would be interesting to test this hypothesis by assessing the distance of those TEs to genes.
4. Figure 2D-F: The authors should indicate the number of TEs that are shown. It would also be relevant to show how the individual TEs behave in the mutants using, e.g., boxplots with jitter.
5. In Figure 4, the authors use H3K9me2 as a proxy for DNA methylation, which requires explanation. Since the other data were generated using DNA methylation in all three contexts (at least this is what I assume), now using H3K9me2 as a proxy for DNA methylation may lead to different conclusions. This is also evident when comparing the numbers of TEs marked by DNA methylation in Figure 1B and those marked by H3K9me2 in Figure 4A, which are different.
6. The authors propose a specific cis determinant for recruiting H3K27me3 to TEs. Have they tested whether there are common cis elements in the H3K27me3-marked TEs and whether these features are potentially missing or mutated in accessions where the TEs are not marked by H3K27me3?
Minor comments:
7. Line 95: The co-occurrence of H3K27me3, H3K9me2, and DNA methylation was also shown in plants. Please see Sato et al., 2021: DOI: https://doi.org/10.7554/eLife.64593.
8. Line 162: H2A.Z is not associated with transcriptional repression. Please revise this sentence accordingly: "H3K27me3 was previously shown to colocalize with other chromatin marks associated with transcriptional repression at protein-coding genes. These include H2A.Z, a histone variant..."
9. Ecotypes should be changed to accessions throughout the manuscript.
10. Line 239: "This phenomenon observed at orthologous TE insertions indicates a change in the epigenetic state - from standard TE-like DNA methylation to H3K27me3,..." The data and the model rather argue that the direction of the switch is opposite; from H3K27me3 to DNA methylation.
11. Line 305: "here we report for the first time the existence of TEs that can be intact and are targeted by Polycomb instead of DNA methylation." I would be careful with this statement, since several studies have demonstrated that TEs are marked by H3K27me3, and whether or not this occurs on intact or truncated TEs is a minor new point not really changing the main conclusion.

Reviewer 2

In this manuscript by Hure et al describes the finding that more-than-expected number of transposable elements are marked by H3K27me3 instead of mC/H3K9me2 in wild type Arabidopsis thaliana. This result challenges the current dogma that mC/H3K9me2-regulated silencing is the primary mechanisms for TE silencing in flowering plants and proposes that more TEs are regulated by PRCs-based mechanisms. The authors showed that newly inserted TEs, which they reproduced by transformation, can be targeted by H3K27me3. Furthermore, they found the transition between H3K27me3-marked and H3K9me2-marked in a subset of TE copies in natural accessions using the publicly available ChIP-seq data in several accessions. They coined this transition "epigenetic switch", which may be confusing name with ON-OFF switch of gene expression. Finally they approached the molecular basis of the epigenetic switch by conducting GWAS using mCG/mCHG/mCHH polymorphism among accessions, based on the fact that mC/H3K9me2 and H3K27me3 are generally mutually exclusive in TEs. They identified several loci that are known to regulate mC. Overall this manuscript describes interesting new findings about the multilayered TE silencing mechanisms. I have several concerns mainly about data representations in the figures.

Specific points

1. Data representation should be improved so that the authors' claims are convincingly supported. For example, in Fig. 2D-G, they claims that H3K27me3 level is changed in the mutants affecting H2AZ and H2Aub. But how significant the changes are is not shown properly. And how these TEs are selected are not described in anywhere. Another example is Fig. 4A-C. The results of clustering are not easy to follow from the metaplots. Probably better to show by boxplots or violin plots instead of metaplots to show the difference between clusters. In addition, what is the relationship between Fig4A and 4B? The venn diagram is based on clustering or other criterias? 107 seems similar to the sum of cluster 7 and 8, but a bit different. In addition, the list of TEs in each cluster should be included as supplementary file as important information for readers.

2. Does the expression of TEs marked by H3K27me3 change in mutants that lose H3K27me3? It is important to understand the significance of the findings of the manuscript.

3. Why are plots in Fig. S5 have distinct 3 groups? There are no intermediate between 3 groups? Or only TEs with distinct pattern of H3K9me2/H3K27me3 are shown and others inbetween are not shown? In that case, the R values are calculated with all TEs or chosen TEs? (If they only used chosen TE for R value, then the strong correlation may be an artifact) These points should be clarified. In addtion, I assume the numbers on top are accession numbers for natural accession, which is confusing. I recommend to put the names of accessions instead.

4. The result in Fig. 5C is based on unpublished results of nanopore sequencing for 130 accessions. The authors should at least make the sequences of analyzed TE, COPIA12D, public to evaluate the reliability of the result.

5. Fig. 5A shows DNA methylation levels in COPIA12D among natural accessions. The result indicates that several natural accessions have no mCG/mCHG/mCHH. However, the main text says "~ always DNA methylated except in Col-0 where it is marked by H3K27me3". Does it mean literally only one accession (Col-0) lose mC? This point should be clarified.

6. It is not easy to see the difference in chromosomal distribution of TEs in Fig. 1E. Should be quantitatively shown as number of TEs in each 10kb or 100kb.

7. Line 162, "H3K27me3 was previously shown to colocalize with other chromatin marks associated with transcriptional repression at protein-coding genes." The papers of the previous

results should be cited properly here.

8. Fig2A shows the number of co-marked by H3K27me3 and H2AZ or H2AK121ub. I assume the red part is only marked by H3K27me3 and green part is co-marked. The legend on top should be mentioned as "H2A.Z & H3K27me3", "H3K27me3-only", and "H2AK121ub & H3K27me3" from left to right. In addition, criteria for co-marked should be mentioned. For example, longer TEs have more chance to have both H3K27me3 and H2AZ/H2AK121ub peaks compared to shorter TEs even if they are randomly localized. Do they allow the "co-localization" in the whole TE region? Or restrict to the local "co-localization".

Minor points
The 3-digit comma are not visible such as in Fig. 1B,C,F.
I assume the "wild type" in Fig. 1A is Col-0, which should be clearly mentioned.
The numbers of TEs are not shown in Fig. 1D.
Line 150 and 231 "epigenomic features" should be more specific such as "H3K9me2 and H3K27me3".
Line 226 H327me3 should be H3K27me3.

Authors' response to reviewers

Authors' response to the referees' comments

We would like to thank the reviewers for their constructive comments and for giving us the opportunity to reflect on certain aspects of our study and improve the manuscript. We now have addressed each suggestion point by point in this letter. This also includes the general editor concerns about "*the statistical analyses (missing significance) and the data availability (i.e. the sequence of the 130 TEs)*" which have now been resolved. The reviewers' comments are below in **bold type** and our response is in regular type. Text amendments are cited here in the response and highlighted in yellow in the manuscript.

Reviewer #1

1. **Line 141: The functional domains should be specified and included as a data table.**

The functional domains present for each class of TEs have now been specified in an UpSet plot (in Additional file 1: Fig.S1F) which complements the Fig.1D. Of note, for Fig.1D, we replaced the violin plot by a boxplot representing the number of domains in each TE class with a statistical test showing non-significant (ns) differences between the four classes.

Besides, the sentence L.146 was completed with the name of the 5 domains:

*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, **Fig.S1F**).*

The legend of Additional file 1, Fig.S1 was also amended:

"(F) UpSet plot depicting the combination of functional domain for the long TEs of each class. GAG, Group-specific antigens or Capsid protein; PR, protease; INT, integrase; RT, reverse-transcriptase; RH, ribonuclease H."

2. **Figure 1B: The authors broadly show DNA methylation, but it would be relevant to specify the type of DNA methylation that overlaps with H3K27me3. This could possibly hint at how the switching or the establishment of H3K27me3 occurs.**

A boxplot showing the levels of DNA methylation in each context for Class 1 and 2 TEs (DNA methylated only and DNA methylated + H3K27me3 marked respectively) has now been included, with the associated p-values (from Wilcoxon-Mann-Withney U-test), in Fig.S1C. Indeed, significant differences in DNA methylation between classes 1 and 2 can be observed in each context.
The text has been modified accordingly L.136 to add these observations:

L.136: 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 (Brinkman 2012, Statham 2012) and more CHH methylation (Fig.S1C).
The legend of Additional file 1, Fig.S1 was also amended:
"(C) Boxplot showing the CG, CHG and CHH methylation levels for each TE in class1 and class 2. Statistical analysis was performed using Wilcoxon-Mann-Whitney U-test."

3. **Figure 1F: The authors find that mainly DNA TEs and Helitrons are marked by H3K27me3. Since both types of TEs are frequently found to contain transcription factor binding sites and are part of promoters, one possible explanation for this observation could be that these TEs are close to genes and function as promoter elements. It would be interesting to test this hypothesis by assessing the distance of those TEs to genes.**

Firstly, we have now performed statistical analysis in Fig.1F using hypergeometric test to prove that H3K27me3-marked TEs are significantly enriched in Helitrons (p-values, visualized as stars on the plot).
We amended the text (L.155) and the legend as follows:

L.155 "*Further analysis of H3K27me3-marked TEs showed that they are **significantly** enriched in Helitrons for all size categories, as well as LTR retrotransposons for long copies*"

*"**(F)** Stacked bar plot shows distribution of TE super-families according to TE length among all TEs or H3K27me3 marked TEs. **Statistical analysis was performed using Hypergeometric test. * = p-value < 0.05; ** = p-value < 0.01; *** = p-value < 0.001; **** = p-value < 0.0001**"*

Next, to test the hypothesis suggested by the reviewer, we produced density plots showing the distance of the different classes of TEs to the closest gene downstream (Fig.S1H); we took into account only TEs that are within 2kb upstream a gene as this interval classically contains gene promoters in Arabidopsis. We found that the H3K27me3-marked TEs (class 3) are closer to genes than are DNA methylated TEs in the 500bp interval upstream of genes. Besides, we found that the H3K27me3 TEs in that 500bp interval mostly comprise Helitrons which supports the suggested hypothesis.

The text has been amended as follows to present the results of these analyses:
*L.157 "Helitrons are frequently found to contain Transcription Factors binding sites and can act as gene promoters/cis-regulatory modules **(Batista et al, 2019, Zhao et al, 2022, Barro-Trastoy and Köhler, 2024)**. Interestingly, the H3K27me3-marked TEs (class 3) are closer to genes than are DNA methylated TEs (Fig. S1H) as also observed recently (Hisanaga et al. 2023), and in particular in the 500bp upstream of promoters; H3K27me3-marked TEs in this interval comprise mostly helitrons which is in support of a role of these TEs as cis-regulatory modules, possibly with a dual activation / silencing function".*

The legend Fig.S1H was also amended:
*"(H) **Top panel:** Bar plot showing the distribution of closest distance from gene for each class of TE. Top: The distribution takes into account the distances, up to 2kb upstream gene, each bin corresponds to 50bp. Statistical analysis was performed using Chi-squared test."**Bottom panel:** Stacked bar plot showing distribution of TE superfamilies for TEs within 400bp upstream of genes.*

4. **Figure 2D-F: The authors should indicate the number of TEs that are shown. It would also be relevant to show how the individual TEs behave in the mutants using, e.g., boxplots with jitter.**

The number of TEs is indicated inside the box, on the top left/right corner.

To show how the individual TEs behave in mutants, box plots have now been generated and shown in Additional file 1: Fig.S2B, D, E, G. We chose not to include jitters as we felt the difference was clear enough to avoid clutter up the figure.

The text (L.192-203) has been modified to include the new figures:

"(Fig.2D, Additional file 1: Fig.S2A-B)" "(Fig.2E, Additional file 1: Fig.S2C-D)" "(Fig.2F left panel and Fig.2G, Additional file 1: Fig.S2E)" "(Fig.2F right panel, Additional file 1: Fig.S2F-G)"

The legend of Fig.S2 was amended as follows:

"(B) Boxplot showing H2A.Z (dark green) and H3K27me3 (red) levels for TEs that lose the marks in arp6 mutant compared to WT."
"(D) Boxplot showing H2AK121Ub (light green) and H3K27me3 (red) levels for TEs that lose the marks in atbmi1a/b/c mutant compared to WT."
"(E) Boxplot showing H3K27me3 (red) levels for TEs marked by H3K27me3 in WT (class 3) that reach higher level of H3K27me3 in ref6 elf6 jmj13."
"(G) Boxplot showing H3K27me3 (red) levels for 'non-marked' TEs in WT (class 4) that gain H3K27me3 in ref6 elf6 jmj13."

5. **In Figure 4, the authors use H3K9me2 as a proxy for DNA methylation, which requires explanation. Since the other data were generated using DNA methylation in all three contexts (at least this is what I assume), now using H3K9me2 as a proxy for DNA methylation may lead to different conclusions. This is also evident when comparing the numbers of TEs marked by DNA methylation in Figure 1B and those marked by H3K9me2 in Figure 4A, which are different.**

We have now provided the justification for this in the main text:

L.258.*"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 (Du and Johnson 2015, Law et al., 2013) but also CG methylation (Rajakumara, Law et al., 2011) and that CG and CHG methylation are also coupled at heterochromatic loci (Du and Johnson, 2015), we reasoned that H3K9me2 variation (presence/absence) between accessions would be a relevant and stringent readout for a strong variation in epigenetic silencing state if any.*

This choice avoided the need to arbitrarily define and use a combination of three thresholds for DNA methylation in all three sequence contexts that would correspond to a bona fide variation in epigenetic silencing state (constitutive heterochromatin versus facultative heterochromatin).

In Figure 1, only CG methylation was used with a threshold of 50% across TE length for defining the "DNA methylated state" (with the rationale that CG and non-CG DNA methylation are coupled at TEs, *Du and Johnson, 2015*); associated DNA methylation in non-CG context was analyzed concomitantly on the classes defined to validate this choice as shown in Fig. S1B and now Fig.S1C as well. We thank the reviewer for pointing out that our strategy was unclear and have now clarified this in the legend of Fig.1B.
"*… 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".*

In Figure 4, the choice to use H3K9me2 as a proxy for heterochromatic DNA methylation was also motivated by the idea to avoid this type of combined representation (threshold for CG methylation and associated non-CG methylation level) for 8 different TE groups (see Fig.4B) and multiple ecotypes (see Fig.S5).

We acknowledge that using H3K9me2 for constitutively heterochromatic TEs is more stringent than using 50% CG methylation threshold but we thought it was preferable to avoid over-estimating the number of TEs that truly change silencing mode in Figure 4.

6. The authors propose a specific cis determinant for recruiting H3K27me3 to TEs. Have they tested whether there are common cis elements in the H3K27me3-marked TEs and whether these features are potentially missing or mutated in accessions where the TEs are not marked by H3K27me3?

We have tested for a given orthologous TE copy that change epigenetic status across accessions whether the H3K27me3 marked sequences are significantly enriched in some cis elements that would be absent in the DNA /H3K9 methylated ones and we did not find any. In addition, we we counted, in H3K27me3-marked sequences versus, DNA methylated sequences, the number of known PRE (Polycomb-response elements)-like motifs that were previously published as being able to recruit PRC2/PRC1 at genes (Berger et al., 2012, Qüesta et al., 2016, Yuan et al., 2016, Xiao et al., 2017, Zhou et al., 2018): likewise, we did not find any significant differences between H3K27me3-marked and DNA methylated sequences. Finally, we also thought to count the number of CG dinucleotides, CHG and CHH trinucleotides and did not find any significant differences between H3K27me3-marked and DNA methylated sequences.
A likely explanation for the lack of detection of enriched cis-motifs in H3K27me3-marked sequences is that some DNA methylated sequences could have the cis-elements needed to recruit H3K27me3 but this potential to recruit H3K27me3 is masked by the presence of DNA methylation. This hypothesis cannot be tested here because the DNA methylation mutants such as *met1* or *ddm1* in Arabidopsis accessions are not available thus H3K27me3 profiles in absence of DNA methylation are unknown. Yet, 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 (*Rougée et al., 2021, Fig. 1G*).

Thus, we can conclude that it is possible that TEs with the potential to recruit H3K27me3 could be enriched in some specific cis elements as we previously observed but this could only be visualized using mutants for DNA methylation. The current absence of detection of motifs significantly enriched suggests that there are not no cis-motifs that are necessary and sufficient to make a TE target of Polycomb pathways instead of DNA methylation. This is in agreement with our GWAS results showing that the determination of TE epigenetic status is multifactorial and involves trans-factors involved in DNA methylation.

We have now included two figures in Additional file 1 (Fig.S5E and Fig.S5F) and amended the text to discuss this:

*L.328:" ... 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 discrete motif sequences that were significantly enriched in H3K27me3 marked sequences 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 (Berger et al., 2012, Qüesta et al., 2016, Yuan et al., 2016, Xiao et al., 2017, Zhou et al., 2018) 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 cannot be tested here because the DNA methylation mutants such as met1 or ddm1 in Arabidopsis accessions are not available thus H3K27me3 profiles in absence of DNA methylation are unknown. Yet , 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.*

Legends of Figure S5 have been amended accordingly to describe the new figures:
S5E: Number of known PRE-like motifs previously shown to recruit PRC2/PRC1 at genes (Berger et al., 2012, Qüesta et al., 2016, Yuan et al., 2016, Xiao et al., 2017, Zhou et al., 2018) in *COPIA12D* sequences either H3K27me3-marked (red) or DNA methylated (black) across accessions .
S5F: Percentage of CG, CHG and CHH di/tri-nucleotides in *COPIA12D* sequences either H3K27me3-marked (red) or DNA methylated (black) across accessions.
.

7. Line 95: The co-occurrence of H3K27me3, H3K9me2, and DNA methylation was also shown in plants. Please see Sato et al., 2021: DOI: https://doi.org/10.7554/eLife.64593.

We thank the reviewer for pointing this oversight: two references have now been added in the text

L. 96 the two marks can co-occur[32-35(Sato et al., 2021, Moreno-Romero et al, 2021)]

8. **Line 162: H2A.Z is not associated with transcriptional repression. Please revise this sentence accordingly: "H3K27me3 was previously shown to colocalize with other chromatin marks associated with transcriptional repression at protein-coding genes. These include H2A.Z, a histone variant…"**

We have now revised the text to nuance this statement and references were added as also requested by Reviewer #2:

L.178: *H3K27me3 was previously shown to colocalize with other chromatin marks that can be associated with a dynamic transcriptional repression at protein-coding genes (Baile and Calonje , 2024). These include H2AZ (Jamge et al., 2023, Dai et al., 2017, Carter et al., 2017, Wang et al., 2018), -a histone variant associated with responsive genes and transcriptional repression when ubiquitinated on Lysine 129 by PRC1 (Gomez-Zambrano et al., 2019) as well as ubiquitination of H2A (H2A/Z ub) (Kralemann et al.,2020, Zhou et al., 2017), which is deposited by the Polycomb repressive complex PRC1*[50] *and confers to genes a silent transcriptional state yet responsiveness to developmental or environmental cues*[15] .

9. **Ecotypes should be changed to accessions throughout the manuscript.**

This has now been amended throughout the manuscript.

10. **Line 239: "This phenomenon observed at orthologous TE insertions indicates a change in the epigenetic state - from standard TE-like DNA methylation to H3K27me3,…" The data and the model rather argue that the direction of the switch is opposite; from H3K27me3 to DNA methylation.**

We agree with the reviewer that our previous writing "from standard TE-like DNA methylation to H3K27me3" was confusing and misleading and we have now amended the text so that we do not imply a directionality of the switch:

L.272"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****"*

We assume that the reviewer, to suggest a directionality of the switch " H3K27me3 to DNA methylation", refers to i) the transgenic TEs which gain H3K27me3 upon neoinsertion to then presumably reach the same epigenetic state as the endogenous TE (DNA methylated) ii) possibly the fact that accessions marked by H3K27me3 are rarer and this may indicate indeed that the H3K27me3 state is less stable , with TE tending to be DNA methylated. However, we think that this does not necessarily exclude the possibility of a switch from H3K27me3 to DNA methylation: for instance, upon decaying, some TE could lose cytosine methylation and gain H3K27me3 as suggested in the discussion and Figure 6B (see below).

In Fig.6, we represented these non-mutually exclusive scenarios but now the possible bi-directionality has been clarified in the text, in the discussion, where the model Fig.6 is described (addition here in bold):

L.387. 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 TEs able to recruit H3K27me3 de novo. Then at some*

*point, upon TE detection by the RNAi machinery leading to 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).

11. Line 305: "here we report for the first time the existence of TEs that can be intact and are targeted by Polycomb instead of DNA methylation." I would be careful with this statement, since several studies have demonstrated that TEs are marked by H3K27me3, and whether or not this occurs on intact or truncated TEs is a minor new point not really changing the main conclusion.

We are now more careful with the statement (now L.351). We have cited additional other studies reporting H3K27me3 presence on TE relics/truncated TEs in WT. Besides, we have now underlined the biological importance of revealing intact (as opposed to truncated/fossil) TE marked by H3K27me3 which have been overlooked in other studies (such as in *Hisanaga et al, 2023)* and why we think it is important; we have amended the text as follows:

L.351 In higher plants, TEs marked by H3K27me3 in wild type have recently gained interest. While many H3K27me3-marked TE relics have been recently described[35] , here we report ~~for the first time~~ 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 Polycomb group proteins 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 have now observed upon loss of H3K27me3 in swn clf mutants and included in Figure 2I as requested by Reviewer 2.

Reviewer #2

1. Data representation should be improved so that the authors' claims are convincingly supported. For example, in Fig. 2D-G, they claim that H3K27me3 level is changed in the mutants affecting H2AZ and H2Aub. But how significant the changes are is not shown properly. And how these TEs are selected are not described in anywhere.

Data representation has been improved and the significance of changes is now shown in new figures S2B, S2D, S2E and S2G as also requested by reviewer #1 (point 4). A Wilcoxon-Mann-Whitney U-test has been performed to show how the changes are significant and p-value has been added on each boxplot as indicated in the legend of amended Figure S2.

How TEs have been selected is now described at the end of the ChIP-seq section in Material & Methods L.447:

To define lists of TEs that are differentially regulated between various mutants (to then analyze the changes of H3K27me3 levels in Figure 2) or between accessions (to then be able to define

different clusters in Figure 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.

Another example is Fig. 4A-C. The results of clustering are not easy to follow from the metaplots. Probably better to show by boxplots or violin plots instead of metaplots to show the difference between clusters. In addition, what is the relationship between Fig4A and 4B? The Venn diagram is based on clustering or other criteria? 107 seems similar to the sum of cluster 7 and 8, but a bit different. In addition, the list of TEs in each cluster should be included as supplementary file as important information for readers.

We have now generated boxplots to replace the metaplots/heatmaps in Fig.4B in order to better visualize the difference between clusters and amended the legend as follows:
"(B) Boxplots 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."
We added the method of clustering in the ChIP-seq section of Material and Methods.

At each boxplot we have now added the number of corresponding TEs.
Indeed, there was a mistake with the number 107 in figure 4A and we thank the reviewer for pointing this out. It has now been corrected by "**117"** which is the sum of cluster 7 (67) and 8 (50).

The text has been amended as follows, because old Figure 4C has now been removed and old Fig.4D has now been replaced by Fig.**4C**
"*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.4**B**).*"

A Supplementary file (Additional file 2) with the list of the TEs per cluster has been added and mentioned in the text:

L.256 "*Using these data, eight clusters of TEs were defined based on the length coverage of H3K27me3 and H3K9me2 (used as a proxy for DNA methylation) profiles (Fig.**4A-B**, **Additional file 2**)"*

2. Does the expression of TEs marked by H3K27me3 change in mutants that lose H3K27me3? It is important to understand the significance of the findings of the manuscript.

We agree that this is an important analysis to make and we have now included it. To show how the individual TEs behave in mutants, box plots showing TE expression in WT and *swn clf* double mutants (which show a complete loss of H3K27me3 as shown in Fig.S2H) have now been generated and shown in **Figure 2I**; a representative screenshot is shown in **Figure 2H**. We reanalyzed raw expression data from Shu et al, 2019 (https://doi.org/10.1002/pld3.100), produced in wild-type and in the double mutant *swn clf* where H3K27me3 marks are lost. We focused on long H3K27me3-marked TEs that are intact (i.e. more likely to have kept their transcription potential and the focus of Figure 2) and observed a transcriptional reactivation for a quarter of them (9/39) in the double mutant compared to the wild-type condition. For the other 30 TEs, transcriptional conditions (Transcription Factor or TF in particular) may not be present

to allow TE expression, in line with previous observations whereby removing silencing epigenetic marks is not sufficient to induce detectable transcription because corresponding TF expression is also needed (see *Zervudacki et al, 2018 for COPIA93 EVD* or *Ito et al., 2011* for *COPIA78 ONSEN* as examples). Thus, we conclude that TEs can be silenced by PcG in the flowering plant Arabidopsis as previously shown in algae and bryophytes (*Hisanaga et al., 2023*).

We added new panels in Figures 2 and S2 to present these new results and amended the corresponding legends as follows:

"(H) Representative browser view of a TE showing H3K27me3 (red) and RNA (blue) in WT and swn clf double mutant."
"(I) Boxplot showing 9 TEs from class3 (TE>3,5kB) significantly up-regulated in swn clf compared to WT. Statistical analysis was performed using Wilcoxon-Mann-Whitney U-test."

In Additional file 1: Fig.S2:
"(H) Representative browser view of 3 TEs showing H3K27me3 (red) and RNA (blue) in WT and swn clf double mutant."

We also amended the text:

L207: *"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 (Shu et al, 2019), in which virtually all H3K27me3 marks are erased in vegetative tissues (Fig.2H, Fig. S2H). Among long and intact TEs, we found that 9 out of 39 were significantly up-regulated upon in double mutant compared to wild-type (Fig.2I , Fig.S2H). As for the remainder, lack of H3K27me3 may not be sufficient to activate the TE, possibly because these TEs need transcriptional conditions , such as Transcription Factor presence, that are not met in swn clf (in line with the observations that loss of epigenetic silencing marks in not always sufficient to activate TE transcription, Zervudacki et al, Ito et al, 2011) . Thus, we conclude that H3K27me3 at TEs can transcriptionally repress their expression in Arabidopsis as shown in algae and bryophytes (Hisanaga et al., 2023)."*

L.354 (discussion): "*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*"

3. Why are plots in Fig. S5 have distinct 3 groups? There are no intermediate between 3 groups? Or only TEs with distinct pattern of H3K9me2/H3K27me3 are shown and others in between are not shown? In that case, the R values are calculated with all TEs or chosen TEs? (If they only used chosen TE for R value, then the strong correlation may be an artifact) These points should be clarified. In addition, I assume the numbers on top are accession numbers for natural accession, which is confusing. I recommend to put the names of accessions instead.

We now have clarified these two points. Firstly, all intermediates are shown: we now have removed the circle to avoid make them look as a group as we agree it was confusing. All TEs are shown so the correlation is valid. Secondly, the numbers on top were indeed accessions

numbers which we agree was confusing; the accessions numbers have now been replaced by accession name in Additional file 1: Fig. S5, and the legend has been amended accordingly:

"(A) Plots showing correlation between H3K27me3 and H3K9me2 values for each TEs in Col-0(6909), Fly2-2(6024), Vinslöv(9057), KBS-Mac-74(1741), Ty-1(5784), Tos-82-387(1254)."

4. **The result in Fig. 5C is based on unpublished results of nanopore sequencing for 130 accessions. The authors should at least make the sequences of analyzed TE, COPIA12D, public to evaluate the reliability of the result.**

We have now generated a fasta file with the sequences of *COPIA12D* (*AT5G28145*) in the 130 accessions which is available as part of Additional file 3 and we amended the text:

L.323 *"To get more insight into the nature of possible cis-determinants, we extracted the sequence of COPIA12D in 130 accessions extracted from their fully assembled genomes (kindly provided by the Nordborg lab, GMI) (**Additional file 3**)."*

5. **Fig. 5A shows DNA methylation levels in COPIA12D among natural accessions. The result indicates that several natural accessions have no mCG/mCHG/mCHH. However, the main text says "~ always DNA methylated except in Col-0 where it is marked by H3K27me3". Does it mean literally only one accession (Col-0) lose mC? This point should be clarified.**

We thank the reviewer for spotting that mistake: this statement was only true for an earlier, former version of the manuscript draft and for the first dozen of accessions we got access to for both DNA methylation and H3K27me3 data; we then got access to more accessions. We now have corrected this and amended the text as follows:

L.298 *"ATCOPIA12D (AT5TE36920) insertion (shown in Fig.4D) is present in all the assembled accessions **~~and is always DNA methylated except in Col-0 where it is marked by H3K27me3~~** 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"*

6. **It is not easy to see the difference in chromosomal distribution of TEs in Fig. 1E. Should be quantitatively shown as number of TEs in each 10kb or 100kb**.

We agree with this and added a circular plot using 100kb bin size in Additional file 1: Fig.S1G to allow better chromosomal visualization of long TEs to complement the Fig.1E**.** We amended the text accordingly:

*"Interestingly, H3K27me3-marked long TEs (>3.5 kb) are dispersed throughout the euchromatic arms and pericentromeric regions (Fig.1A, Fig.1E **and Additional file 1: Fig.S1G**), in contrast to the DNA-methylated ones, which are generally clustered in pericentromeres."*

As well as the legend:

"(G) Circos plot showing chromosomal distribution of TEs>3.5kb for each class from class 1 to class 4 (outside to inside, respectively) using 100kb bin size."

7. **Line 162, "H3K27me3 was previously shown to colocalize with other chromatin marks associated with transcriptional repression at protein-coding genes." The papers of the previous results should be cited properly here.**

We now added references for the previous results and amended the text, as also suggested by reviewer #1:

L.178: *H3K27me3 was previously shown to colocalize with other chromatin marks that can be associated with a dynamic transcriptional repression at protein-coding genes (Baile and Calonje, 2024). These include H2AZ (Jamge et al., 2023, Dai 2017, Carter 2017, Weng et al., 2018_ref 43-46_), -a histone variant associated with responsive genes and transcriptional repression when ubiquitinated on Lysine 129 by PRC1 (Gomez-Zambrano et al., 2019_ref.47_) as well as H2A/Z ub (Kralemann et al., 2020, Zhou et al., 2017_ref.48-49_), which is deposited by the Polycomb repressive complex PRC1[50] and confers to genes a silent transcriptional state yet responsiveness to developmental or environmental cues[15].*

8. **Fig2A shows the number of co-marked by H3K27me3 and H2AZ or H2AK121ub. I assume the red part is only marked by H3K27me3 and green part is co-marked. The legend on top should be mentioned as "H2A.Z & H3K27me3", "H3K27me3-only", and "H2AK121ub & H3K27me3" from left to right. In addition, criteria for co-marked should be mentioned. For example, longer TEs have more chance to have both H3K27me3 and H2AZ/H2AK121ub peaks compared to shorter TEs even if they are randomly localized. Do they allow the "co-localization" in the whole TE region? Or restrict to the local "co-localization".**

The panel in Fig.2A has been changed according to the reviewer's suggestion (legend on top) for clarity. The overlap was done based on peaks overlap which we have now clarified in the legend, as follows:

"(A) Stacked bar plot showing distribution of TE marked by, either 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."***

Minor points:

- **The 3-digit comma are not visible such as in Fig. 1B,C,F.**

Space was now introduced between digits.

- **I assume the "wild type" in Fig. 1A is Col-0, which should be clearly mentioned.**

This has now been clarified in the legend.

- **The numbers of TEs are not shown in Fig. 1D.**

A new figure was made and the number of TE for each class was added.

- **Line 150 and 231 "epigenomic features" should be more specific such as "H3K9me2 and H3K27me3".**

The text has been amended as suggested.

- **Line 226 H327me3 should beH3K27me3.**

The text has been modified.

Second round of review

Reviewer 1

The authors have successfully addressed my concerns and improved the manuscript. I suggest to rephrase one statement and to correct some typos, otherwise I have no further concerns.
Line 97: remove wild type
Line 109: remove abbreviation for TE
Line 187: add comma before which
Line 215: remove comma
Line 216 Close the bracket
Line 212: remove upon
L331: sequence motifs
L332: remove marked sequences
Line 339: "This hypothesis cannot be tested here because the DNA methylation mutants such as met1 or ddm1 in Arabidopsis accessions are not available"; this is not correct; there are met1 mutants published in 18 accessions: Srikant et al., Genome Biology 2022.
L342: ddm1 in italics
L357: use abbreviation for Polycomb-group proteins

Reviewer 2

The authors appropriately revised the manuscript according to my comments on the original manuscript. I have some more comments on the revised manuscript below.

1. I'm still not convinced that the Fig.S5A shows all the TEs but not chosen TEs. First of all, the numbers of TEs shown is ~14,000 for Col-0, which is much fewer than "Total TEs 31,060" in Fig. 1B. In addition, the discontinuous (zero or more than ~1.2) levels of H3K9me2 and H3K27me3 in the scatter plots are not typical patterns of ChIP-seq. The authors should clarify this point in the legend or text.
2. In Fig.2D~F, how these TEs are chosen is not described. For example in Fig. 2D, is it based on H2A.Z or on H2A.Z AND H3K27me3? This is important point to interpret the results. Furthermore, assuming from the numbers of TEs analyzed, the authors limited the analysis to the long TEs (>3.5kb). I think that analyzing more TEs in different size ranges will increase the power of analysis, which will strengthen (or reject) their conclusion.
3. Related to the above comment 2, newly added RNA analyses should also be more general. They only showed "upregulated" TEs (n=9), which I think could be misleading. How general the trend is should be shown.
4. In response to my previous minor comment about "wild type" in the legend of Fig. 1A, no fix is made. I recommend to mention "Col-0" rather than wild-type (WT) in the legend of Fig. 1A. Or like "wild-type Col-0".

5. Newly added circular plot (Fig. S1G) is good, but the inner plot (class4) is not visible. It may because the range is too large (0-7) or because of the color.

Authors' response to reviewers

Responses to Reviewers' comments

We thank the reviewers for their time and careful reading of the revised manuscript which enabled us to improve it; we now have addressed all the concerns raised. Below are the reviewers comments in bold and our detailed response.

We also want to notify the editor that we added a table (Additional File 2, Table S1) that contains the list of TEs of each class presented in Figure 1 as it could be useful for the community.

Reviewer #1: The authors have successfully addressed my concerns and improved the manuscript. I suggest to rephrase one statement and to correct some typos, otherwise I have no further concerns.

Line 97: remove wild type

This has been fixed.
Line 109: remove abbreviation for TE

This has been fixed.
Line 187: add comma before which

This has been fixed.
Line 215: remove comma

This has been fixed.
Line 216 Close the bracket

This has been fixed.
Line 212: remove upon

This has been fixed.
L331: sequence motifs

This has been fixed.
L332: remove marked sequences

This has been fixed.
Line 339: "This hypothesis cannot be tested here because the DNA methylation mutants such as met1 or ddm1 in Arabidopsis accessions are not available"; this is not correct; there are met1 mutants published in 18 accessions: Srikant et al., Genome Biology 2022.

We now have corrected the text and added that reference (68):

" this hypothesis could be tested in the future by profiling H3K27me3 marks in DNA methylation mutants such as met1 or ddm1 in Arabidopsis accessions [68]".
L342: ddm1 in italics

This has been fixed.
L357: use abbreviation for Polycomb-group proteins
This has been fixed.

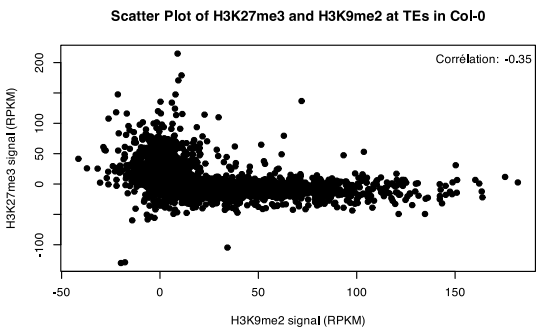
Reviewer #2: The authors appropriately revised the manuscript according to my comments on the original manuscript. I have some more comments on the revised manuscript below.

1. I'm still not convinced that the Fig.S5A shows all the TEs but not chosen TEs. First of all, the numbers of TEs shown is ~14,000 for Col-0, which is much fewer than "Total TEs 31,060" in Fig. 1B. In addition, the discontinuous (zero or more than ~1.2) levels of H3K9me2 and H3K27me3 in the scatter plots are not typical patterns of ChIP-seq. The authors should clarify this point in the legend or text.

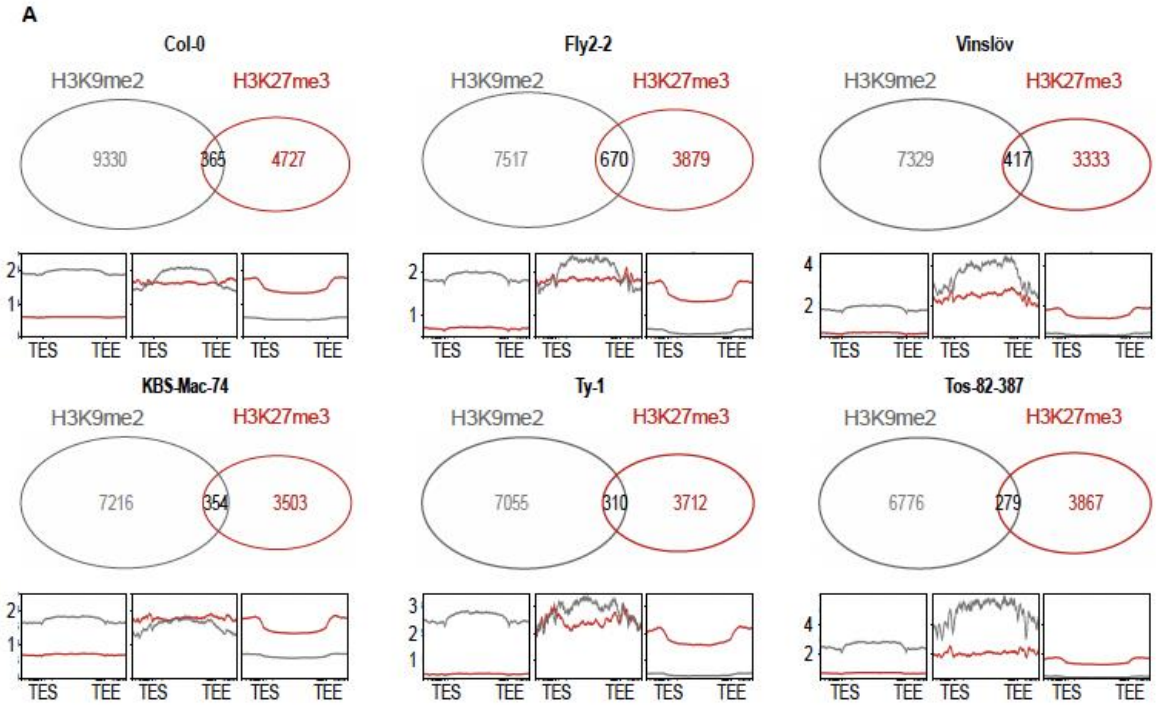
- We agree that we did not clarify this enough in the previous response. In fact, only TEs marked by either H3K9me2 or H3K27me3, which are relevant for this analysis are shown, and not the TEs that are marked by none of the marks. In Col accession, there are about 12,000 "non-marked TEs" (see Class4 of Fig.1) which explains the discrepancy in the numbers between Fig.1 and Fig. S5 for the most part. Secondly, we used in this Figure S5A H3K27me3 and H3K9me2 profiles from Kornienko et al., 2023. As the efficiency of the ChIPs in this dataset is not exactly the same as in our dataset used in Fig.1B, this can explain that there are less peaks found by peak calling, thus explaining the remainder of the differences as for the TE number.

We now have amended the text and the figure legend accordingly to make this more clear to the reader (see further below).

-As for the discontinuous values, we realized that this is due to the fact that for the TEs with a peak (and value) for H3K27me3 and no peak for H3K9me2, there was no available score for H3K9me2 in our IDR-MACS2 ChIP-seq analysis output, thus we replaced the absence of value by a zero (and vice versa for TEs marked by H3K9me2 with no peaks for H3K27me3). We acknowledge that this could be confusing and introduces a bit of inaccuracy. Thus, we first thought to use the same lists of TEs based on peak detection for each mark, but attributed a score based on RPKM. An exemple of such new scatterplot is shown below: while the anti-correlation is less strong (due to the low RPKM values), the essential result is the same i.e. most TEs that have a H3K27me3 peak have low RPKM score for H3K9me2 and most TE that have a H3K9me2 peak have a low RPKM score for H3K27me3.



The reviewer's comment made us realize that this representation is not the best for our purposes, because our point is to show, that, in all ecotypes tested, **most TEs which have a H3K27me3 peak do not have a H3K9me2 peak and vice-versa (to be able to proceed with GWAS)**. In order to show this point more directly and in a clearer manner to the reader, we now have decided to replace the scatterplots by a combination of a venn-diagrams (TEs with a H3K27me3 peak and TEs with a H3K9me2 peak) and three metaplots in each ecotype which are shown in Fig.S5A.



We amended the legend Fig.S5 accordingly:

(A) ***Venn diagram*** *showing in 6 different accessions (Col-0 (6909), Fly2-2(6024), Vinslöv(9057), KBS-Mac-74(1741), Ty-1(5784),Tos-82-387(1254)), the number of TEs marked by either H3K9me2 (as a proxy for DNA methylation - left), or by H3K27me3 (right) or by both (middle),* ***using data from Kornienko et al., 2023****.* ***TE with no peak detected for none of the marks are not shown.*** *For each subset, associated average levels of H3K9me2 (grey) and H3K27me3 (red) are shown below in the metaplots.*

As well as the main text:

L.289: *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[60].

2. In Fig.2D~F, how these TEs are chosen is not described. For example in Fig. 2D, is it based on H2A.Z or on H2A.Z AND H3K27me3? This is important point to interpret the results.

In that figure, the TEs used were the ones for which we detected a loss of peak or a gain of peak in the mutants tested. Thus we used only the subset that was impacted specifically by the mutations to address the extent of the epigenetic changes and avoid diluting/masking the effects with TEs which are not dependent on the pathways tested.

We now have amended the legends of Fig.2D-E accordingly to make this point clearer to the reader:

:…"*at the TE subset impacted by the mutation*" is added in each relevant panel.

For Fig.2F, this is already explicited in the legend.

Furthermore, assuming from the numbers of TEs analyzed, the authors limited the analysis to the long TEs (>3.5kb). I think that analyzing more TEs in different size ranges will increase the power of analysis, which will strengthen (or reject) their conclusion.

It is right that we initially presented the set of TEs >3,5kb where the three marks co-localize the most frequently. But we agree with the reviewer that including more TEs in different size ranges increase the power of analysis and we now have included the requested analysis in Figure S2B,D,E,G. The statistical significance is indeed increased, strengthening our initial conclusion.

Besides, the metagenes of Figures 2D-F (initially made using TEs >3,5 kb) have now been replaced by metagenes using TEs of all size: the profile obtained is now smoother and leads to the same conclusion.

The legend has been amended accordingly and the text as well (no longer restricting the conclusion to TEs>3,5 kb)

L.205: *These results demonstrate that TEs can harbor chromatin hallmarks of typical PcG genic targets, including active regulation by histone demethylase.*

3. Related to the above comment 2, newly added RNA analyses should also be more general. They only showed "upregulated" TEs (n=9), which I think could be misleading. How general the trend is should be shown.

We have now have repeated the RNA seq analyses on all H3K27me3-marked TEs of all sizes: we find that 3161 out 4656 show significant upregulation in *swn-clf* , which we plotted on Fig. 2I. Indeed, making this analysis more general, strengthened its power.

We amended the text accordingly:

L.211: *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).*

4. In response to my previous minor comment about "wild type" in the legend of Fig. 1A, no fix is made. I recommend to mention "Col-0" rather than wild-type (WT) in the legend of Fig. 1A. Or like "wild-type Col-0".

This is has now been corrected.

5. Newly added circular plot (Fig. S1G) is good, but the inner plot (class4) is not visible. It may because the range is too large (0-7) or because of the color.

We changed the colour to blue and increased the size of the panel a little.

Evaluation. Has the author satisfactorily responded to your
previous review?</br>

Reviewer #1: Yes

Reviewer #2: Yes

Third round of review

Reviewer 2

The authors responded to my previous comments appropriately. There is no further requests from me. Congratulations on this new piece of work describing novel connections between DNA methylation- and PcG-mediated TE silencing!