THE EMBO JOURNAL

Manuscript EMBO-2016-94983

Gene-body chromatin modification dynamics mediate epigenome differentiation in Arabidopsis

Soichi Inagaki, Mayumi Takahashi, Aoi Hosaka, Tasuku Ito, Atsushi Toyoda, Asao Fujiyama, Yoshiaki Tarutani and Tetsuji Kakutani

Corresponding author: Tetsuji Kakutani, National Institute of Genetics

Review timeline:

Submission date:	09 June 2016
Editorial Decision:	06 July 2016
Revision received:	04 October 2016
Editorial Decision:	01 November 2016
Revision received:	14 November 2016
Editorial Decision:	07 December 2016
Revision received:	14 December 2016
Accepted:	16 December 2016

Editor: Anne Nielsen

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 06 July 2016

Thank you for submitting your manuscript for consideration by the EMBO Journal. It has now been seen by three referees whose comments are shown below.

As you will see from the reports, our three referees all express great interest in the findings reported in your manuscript, although they also raise a number of technical and conceptual concerns that you will have to address in full before they can support publication in The EMBO Journal. You will see that the referees generally agree that your findings on the crosstalk between H3K9me2 and H3K4me3 on gene bodies is intriguing but that the current manuscript falls short in sufficiently analysing the available data and discussing it in the context of your previously published work. In addition, you will see that all three referees find that the manuscript need to be extended, both in writing the introduction and at the level of experimental downstream analysis. We realize that these are challenging experiments of a potentially uncertain outcome but if you would undertake the effort to revise the manuscript as requested by the three referees we would be happy to look at a revised manuscript. Given the need for extensive revision we could also offer to extend the deadline for submitting the revised manuscript to 6 months from now (rather than the usual 3).

Based on the overall interest and the constructive comments from the referees, I would thus like to invite you to submit a revised version of the manuscript, addressing the comments of all three reviewers. I should add that it is EMBO Journal policy to allow only a single round of revision, and acceptance or rejection of your manuscript will therefore depend on the completeness of your responses in this revised version. In addition, I would ask you to provide us with a preliminary

point-by-point response at this stage already, outlining the additional experiments and analysis that you would be able to include in a revised manuscript.

When preparing your letter of response to the referees' comments, please bear in mind that this will form part of the Review Process File, and will therefore be available online to the community. For more details on our Transparent Editorial Process, please visit our website: http://emboj.embopress.org/about#Transparent_Process

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

REFEREE REPORTS

Referee #1:

In the manuscript entitled "The gene-body chromatin modification dynamics mediates epigenome differentiation in Arabidopsis" Inagaki et al identify a mutation in the histone demethylase LDL2 suppresses the developmental phenotype associated with hyper H3K9me2 accumulation in ibm1 mutants. The ibm1 ldl2 double mutant suppresses the growth phenotype, but does not alter the H3K9me2 accumulation seen in ibm1. This phenotype suppression in the double mutant is found to act through the retention of H3K4me1 in gene bodies, this tying together the dynamic regulation of two chromatin marks: H3K9me and H3K4me1. RNA-seq gene expression analysis was performed and a model was proposed for gene regulation through this dynamic chromatin regulation. Lastly, the suvh456 triple mutant was investigated, as it has very low levels of H3K9me, and it was demonstrated that H3K9me and H3K4me1 seem to be mutually exclusive marks the are coordinated: when one goes up, the other goes down, and vice versa.

I found the overall conclusions of this manuscript very interesting, but I also found that some of the conclusions were not supported by the data. Therefore, in my eyes this publication currently does not meet the standards of EMBO Journal. I have listed these inconsistencies below.

1. In Figure 1B I noticed that the number of "genes" analyzed was very high for Arabidopsis (33,602). The accepted number of genes for this organism is ~27,000. This makes me believe that transposable elements (TEs) are littering this "gene" database, as many annotated "genes" in Arabidopsis overlap TEs. When TEs are separated and investigated (such as in Figure S8), the largest change in H3K4me1 and H3K9me is observed. Therefore, I wonder how much the H3K9me/H3K4me1 dynamics are actually occurring on real (non-TE) genes? This is important, as throughout the paper, and in the manuscript title, "gene body" dynamics are referred to, but I'm not convinced that these are real gene body dynamics and not TE-body dynamics. An analysis of clean genes (not overlapping the TE annotation) should be performed.

2. On page 4 the authors write that the Type II mutants "suppressed the ibm1-induced developmental phenotypes with keeping the ectopic DNA methylation", and then reference several figures. However, in none of those figures, or throughout the entire manuscript, is DNA methylation analyzed. Thus, this claim is unsupported.

3. On page 6 the authors write that down regulated (DG) genes showed ectopic accumulation of H3K9me, but most of the up regulated genes (UGs) did not. The authors then go on to explain how the DG and UG sets are thus different (direct vs. indirect targets of LDL2). However, upon inspection of Figure 1G, it is quite clear that both UG and DG gene sets do have increased levels of H3K9me, and thus their claim and theory are unsupported.

4. In addition to the UG and DG sets, it seems that there are many genes for which expression does not go back to wild-type levels in the double mutant. These genes are never discussed but important. Is there any explanation for their regulation?

5. Do the TEs that lose H3K9me but do not gain H3K4me1 become expressed? This is an important

question as the answer will demonstrate if H3K4me1 changes are required for expression, a consequence of expression, or independent of expression. This point may have been made in the manuscript text, but I was not able to follow.

6. Please include references to the data availability and access for the large datasets. Also, the number of replicates is never shown and bioinformatic analysis section of the Methods is underdeveloped and needs to be expanded.

7. The Introduction section of this manuscript is too short and does not properly prepare the reader for the rest of the manuscript. The Introduction reads as a second Abstract, and key aspects of the first section of Results (such as ectopic DNA methylation and the developmental defects of ibm1) are never touched upon.

8. The two parts of the paper, first dealing with LDL2, and the second with SUVH456, seem very disjointed. Can a mutant combination between the two be created or is there any other way to connect the two halves of the manuscript?

9. It would be helpful to include an analysis of the ldl2 single mutant. It is unclear what this protein is doing in a wild-type cell, since it is only analyzed in a double mutant background with a second mutation that has strong molecular and phenotypic consequences.

Referee #2:

The cross-talk between different histone modifications is an important topic in the epigenetics field although the molecular mechanism underlying this phenomenon is still largely unknown. Previous papers from the same lab showed that mutations of IBM1, which encodes a H3K9me2 demethylase, can induce ectopic H3K9me2 in genic regions of Arabidopsis on a genome-wide basis. The ibm1 mutant shows severe developmental abnormality. Through a forward genetic screen, Inagaki et al. have now obtained ldl2 mutation which reverses the ibm1 developmental defect but has no effect on H3K9me2 patterns of ibm1. These results suggest that the ldl2 mutation recovers ibm1 developmental defects downstream of H3K9me2. Interestingly, the author found increase of H3K4me1 levels in a subset of genes in ibm1ldl2 double mutant compared to ibm1 single mutant. Based on these results they defined a cross-talk between two histone modifications: H3K9me2 and H3K4me1, in intragenic regions of the Arabidopsis genome. These observations are interesting and of scientific significance but the molecular mechanism remains unknown. Several issues need to be addressed before this work can be further considered.

Major concerns
1), In previous reports, the authors claimed there is no correlation between ectopic gene body H3K9me2 induced by the ibm1 mutation and transcription. For the marker gene BONSAI, although intragenic CHG DNA methylation and H3K9me2 increase in the ibm1 mutant background, BONSAI transcript levels remain unchanged. In this manuscript, the authors concluded that down-regulated genes (DGs) may be the direct targets of IBM1 whereas the up-regulated genes (UGs) may be due to indirect effects, although there are more UGs in ibm1 and the number of recovered UGs by further ldl2 mutation is significantly higher than that of DGs. In addition, the authors previously showed that the major targets of IBM1 are transcribed, low-copy protein coding genes (that may be the reason why the ibm1 mutant shows severely developmental defect) whereas in this manuscript the majority of the claimed direct targets of IBM1 are transposon genes. These apparent discrepancies in results generated by the same group at different times need to be addressed and clarified. To identify the direct targets of IBM1, we feel it is necessary to perform a ChIP-seq. experiment to map the genomic occupancy of IBM1. The data would help readers to understand the biological function of IBM1 and illuminate the significance of the cross-talk between H3K9me2 and H3K4me1.

2), Although the author showed that the H3K4me1 levels decreased in a subset of genes in ibm1ldl2 double mutant compared to the imb1 single mutant, it was disappointing that no attempt was made to address the biological function of LDL2. More specifically is the LDL2 a bona fide H3K4me1 demethylase? Whether the K436A mutation indeed compromised the presumed LDL2 demethylase activity is questionable. There could be trivial explanations; the mutant phenotype could be caused

by a change in protein stability and/or cellular localization. At least three experiments should be conducted to address these issues, a) In vitro/in vivo assay to detect the demethylase activity of LDL2, b) ChIP-Seq. to compare H3K4me1 levels between ldl2 mutant and WT. c) Western blot to analyse possible changes in the LDL2 (WT) and LDL2(K436A) protein levels in transgenic plants (Fig. S1E).

Minor concerns,
1), The use of abbreviations e.g. UGs/UTGs and DGs/DTGs has to be consistent between figure legend and labelling in the figure (cf. Fig 1G).

2), The INTRODUCTION is too short and should be expanded. The authors have published several papers on IBM1 and not all readers are familiar with their considerable historical data. The authors should provide more background in the INTRODUCTION section.

3), All data shown in the manuscript were not tested for statistical significance. This should be rectified.

Referee #3:

Histone H3 lysine 9 methylation (H3K9me) is a repressive epigenetic mark that is found not only at promoter regions of transcriptionally silent genes but also on internal regions (gene bodies), where its biological significance is not well understood. As the authors have previously shown, mutations in IBM1, a histone H3 lysine 9 (H3K9) demethylase, result in ectopic H3K9me2 methylation at gene bodies and severe developmental defects in Arabidopsis. In the present study, they carried out a forward genetic screen on the ibm1 mutant to identify mutations that suppress the developmental defects in ibm1. This screen identified LDL2, a putative H3K4 demethylase as a mediator of the ibm1-induced developmental defects.
Further work using mRNA-seq and ChIP-seq to examine transcription and histone modifications on a genome-wide level in various mutant combinations led the authors to propose a model for differentiation of bi-stable epigenetic states at gene bodies through the activities of two histone demethylases: transcription induces loss of H3K9me2 through the action of IBM1, and H3K9me2 induces loss of H3K4me through the action of LDL2.

This study provides new and interesting information on how dynamic changes in histone modifications in gene bodies influence transcription. There are four LDL genes in the Arabidopsis genome and currently only functional information about one (LDL4 or FLD). This study adds new information on this important, conserved group of histone methyltransferases and will be interesting for epigeneticists working not only on plants but other systems as well.

1. Mention the name of the genes responsible for Type I mutations (CMT3 and SUVH4) in Results, not just in Materials and Methods.

2. The paper is quite short (only 3 figures in the main text). I recommend moving some important supplementary information into the main text. For example, show in the main text Venn diagrams currently in Fig S2, S3, Fig, 5A histone H3 methylation patterns in regions around DGs.

3. The experiments with suvh456 triple mutants, which are defective in three H3K9 methyltransferases that act redundantly to silence transposable elements, seem tacked on (e.g they are not mentioned in the Abstract). More effort could be made to integrate these results into the whole 'story' by adding something to the Abstract and Perspective section. Some results from the suvh456 experiments - e.g. Fig. S7 part A and Fig. S8 part A - could be shown in the main text.

3. The acronyms - DGs, UGs, DEGs, TEGs - get confusing. Except for DEGs, write out 'down-regulated' and 'up-regulated'. Refer to TEGs as 'TE genes'.

4. Although the authors aim 'To understand the molecular basis of the LDL2-mediated developmental defects' (p. 5), aside from identifying numbers of differentially expressed genes by mRNA-seq, there is no further insight into the developmental aspects purportedly being addressed in

the study. Why are developmental defects of imb1 mutant suppressed in imb1 ldl2 if ectopic H3K9me is not removed in this double mutant? What is basis of development defects in ibm1? Further discussion of these points is needed; perhaps a GO analysis of the DEGs would help to shed some light on these questions.

1st Revision - authors' response 04 October 2016

Point-by-point response to the referee's comments.
(The referee's comments and our responses are shown by black and blue, respectively.)

Referee #1

We are glad to hear that this referee regards "the overall conclusions of this manuscript very interesting". We also thank the constructive comments. We addressed every point raised by this referee.

In the manuscript entitled "The gene-body chromatin modification dynamics mediates epigenome differentiation in Arabidopsis" Inagaki et al identify a mutation in the histone demethylase LDL2 suppresses the developmental phenotype associated with hyper H3K9me2 accumulation in ibm1 mutants. The ibm1 ldl2 double mutant suppresses the growth phenotype, but does not alter the H3K9me2 accumulation seen in ibm1. This phenotype suppression in the double mutant is found to act through the retention of H3K4me1 in gene bodies, this tying together the dynamic regulation of two chromatin marks: H3K9me and H3K4me1. RNA-seq gene expression analysis was performed and a model was proposed for gene regulation through this dynamic chromatin regulation. Lastly, the suvh456 triple mutant was investigated, as it has very low levels of H3K9me, and it was demonstrated that H3K9me and H3K4me1 seem to be mutually exclusive marks the are coordinated: when one goes up, the other goes down, and vice versa.

I found the overall conclusions of this manuscript very interesting, but I also found that some of the conclusions were not supported by the data. Therefore, in my eyes this publication currently does not meet the standards of EMBO Journal. I have listed these inconsistencies below.

1. In Figure 1B I noticed that the number of "genes" analyzed was very high for Arabidopsis (33,602). The accepted number of genes for this organism is ~27,000. This makes me believe that transposable elements (TEs) are littering this "gene" database, as many annotated "genes" in Arabidopsis overlap TEs. When TEs are separated and investigated (such as in Figure S8), the largest change in H3K4me1 and H3K9me is observed. Therefore, I wonder how much the H3K9me/H3K4me1 dynamics are actually occurring on real (non-TE) genes? This is important, as throughout the paper, and in the manuscript title, "gene body" dynamics are referred to, but I'm not convinced that these are real gene body dynamics and not TE-body dynamics. An analysis of clean genes (not overlapping the TE annotation) should be performed.

Our response: It is true that the 33,602 "genes" include TE genes. Importantly, however, change in expression by *ibm1* was mainly found in cellular genes. Within 443 DGs, 432 were cellular genes, 3 were TE genes, 6 were non-coding genes, and 2 were pseudogenes. Within 1220 UGs, 1155 were

cellular genes, 47 were TE genes, 5 were non-coding genes, and 13 were pseudogenes. We agree with the referee that these are important points and we included the information above in the revised manuscript. In addition, in order to make distinction between cellular genes and TE genes clearer, we used different colors for each of them in Figures such as 2(A), 3, 5, and 6. Through these improvements, it became clearer that genes, rather than TEs, are major targets of IBM1.

2. On page 4 the authors write that the Type II mutants "suppressed the ibm1-induced developmental phenotypes with keeping the ectopic DNA methylation", and then reference several figures. However, in none of those figures, or throughout the entire manuscript, is DNA methylation analyzed. Thus, this claim is unsupported.

Our response: In the original version, DNA methylation status of the Type II mutants has been described briefly in Materials and Methods, from the bottom paragraph of page 11 to page 12. In the revised manuscript, we added results of whole genome bisulfite sequencing of *ibm1-ldl2* and sibling *ibm1-LDL2* plants, because DNA methylation status is an important part of the story. The results show that ectopic DNA methylation of *ibm1* does not recover in *ibm1-ldl2* mutant, even in genes whose expressions are decreased in *ibm1* and recovered in *ibm1 ldl2* (new Fig 2E). That is consistent with our ChIP-seq results for H3K9me2 (new Figure 2A-C) and further support our conclusion that LDL2 functions downstream of DNA/H3K9 methylation.

3. On page 6 the authors write that down regulated (DG) genes showed ectopic accumulation of H3K9me, but most of the up regulated genes (UGs) did not. The authors then go on to explain how the DG and UG sets are thus different (direct vs. indirect targets of LDL2). However, upon inspection of Figure 1G, it is quite clear that both UG and DG gene sets do have increased levels of H3K9me, and thus their claim and theory are unsupported.

Our response: Large proportion of genes (~4,000) accumulates ectopic H3K9me2 in *ibm1*. What we meant in this part was that increase of H3K9me2 in UGs is not overrepresented than that in the total genes (actually they are much underrepresented; new Fig 2CD). In the revised manuscript, we clarified these points, with adding statistical tests in the legend of Figure 2.

4. In addition to the UG and DG sets, it seems that there are many genes for which expression does not go back to wild-type levels in the double mutant. These genes are never discussed but important. Is there any explanation for their regulation?

Our response: That is true that not all UG and DG go back to wild-type level in *ibm1-ldl2* double mutant. Interestingly, GO analyses revealed that UGs as well as genes up-regulated in *ibm1* and not fully recovered in *ibm1 ldl2* are enriched in defense-related genes. We believe the results convincing, as they are reproducibly seen in triplicate samples. We briefly discussed that at the beginning of Perspective part (p.11).

5. Do the TEs that lose H3K9me but do not gain H3K4me1 become expressed? This is an important

question as the answer will demonstrate if H3K4me1 changes are required for expression, a consequence of expression, or independent of expression. This point may have been made in the manuscript text, but I was not able to follow.

Our response: We thank the referee for the very constructive suggestion. We analyzed relationship between H3K4me1 and transcriptional de-repression in TEs loosing H3K9me2 in *suvh456* (shown in new Figure 6E). Very interestingly, almost all of TEs without gain of H3K4me1 remain silent, further supporting our interpretation that H3K4me1 functions upstream of transcriptional de-repression.

6. Please include references to the data availability and access for the large datasets. Also, the number of replicates is never shown and bioinformatic analysis section of the Methods is underdeveloped and needs to be expanded.

Our response: We deposited the sequence data to a database and added the accession number in the manuscript in the Materials and Methods section. Number of replicates in each experiment is added in the Materials and Methods section. We also added more detailed information of the bioinformatic analyses in the Materials and Methods.

7. The Introduction section of this manuscript is too short and does not properly prepare the reader for the rest of the manuscript. The Introduction reads as a second Abstract, and key aspects of the first section of Results (such as ectopic DNA methylation and the developmental defects of ibm1) are never touched upon.

Our response: As suggested, we described the background and key aspects, such as properties of ectopic DNA methylation and H3K9me2 induced by *ibm1*, more thoroughly in the Introduction.

8. The two parts of the paper, first dealing with LDL2, and the second with SUVH456, seem very disjointed. Can a mutant combination between the two be created or is there any other way to connect the two halves of the manuscript?

Our response: We added explanation about the relationship of results in these two parts, so that the connection becomes clearer.

9. It would be helpful to include an analysis of the ldl2 single mutant. It is unclear what this protein is doing in a wild-type cell, since it is only analyzed in a double mutant background with a second mutation that has strong molecular and phenotypic consequences.

Our response: We agree that analyses of ldl2 single mutant are important. We added ChiP-seq and RNA-seq results of the *ldl2* single mutant and control WT plants. Although the effect of *ldl2* mutation is not large as in *ibm1* mutant background, some TEs are affected in both H3K4me1 and

transcription, supporting the idea that targets of LDL2 are transcription units with H3K9me2.

Referee #2:

We are glad to hear that this referee regards our manuscript as "These observations are interesting and of scientific significance". We also thank the referee for constructive comments. We addressed most of the points raised by this referee.

The cross-talk between different histone modifications is an important topic in the epigenetics field although the molecular mechanism underlying this phenomenon is still largely unknown. Previous papers from the same lab showed that mutations of IBM1, which encodes a H3K9me2 demethylase, can induce ectopic H3K9me2 in genic regions of Arabidopsis on a genome-wide basis. The ibm1 mutant shows severe developmental abnormality. Through a forward genetic screen, Inagaki et al. have now obtained ldl2 mutation which reverses the ibm1 developmental defect but has no effect on H3K9me2 patterns of ibm1. These results suggest that the ldl2 mutation recovers ibm1 developmental defects downstream of H3K9me2. Interestingly, the author found increase of H3K4me1 levels in a subset of genes in ibm1ldl2 double mutant compared to ibm1 single mutant. Based on these results they defined a cross-talk between two histone modifications: H3K9me2 and H3K4me1, in intragenic regions of the Arabidopsis genome. These observations are interesting and of scientific significance but the molecular mechanism remains unknown. Several issues need to be addressed before this work can be further considered.

Major concerns

1), In previous reports, the authors claimed there is no correlation between ectopic gene body H3K9me2 induced by the ibm1 mutation and transcription. For the marker gene BONSAI, although intragenic CHG DNA methylation and H3K9me2 increase in the ibm1 mutant background, BONSAI transcript levels remain unchanged. In this manuscript, the authors concluded that down-regulated genes (DGs) may be the direct targets of IBM1 whereas the up-regulated genes (UGs) may be due to indirect effects, although there are more UGs in ibm1 and the number of recovered UGs by further ldl2 mutation is significantly higher than that of DGs. In addition, the authors previously showed that the major targets of IBM1 are transcribed, low-copy protein coding genes (that may be the reason why the ibm1 mutant shows severely developmental defect) whereas in this manuscript the majority of the claimed direct targets of IBM1 are transposon genes. These apparent discrepancies in results generated by the same group at different times need to be addressed and clarified. To identify the direct targets of IBM1, we feel it is necessary to perform a ChIP-seq. experiment to map the genomic occupancy of IBM1. The data would help readers to understand the biological function of IBM1 and illuminate the significance of the cross-talk between H3K9me2 and H3K4me1.

Our response: We agree with the referee that Chip-seq of IBM1 is informative. We added the results in the revised manuscript. The results show that IBM1 is localized within the body of expressed

genes but not in non-expressed genes or TE genes (Fig 2EF). The results further support our model that IBM1 removes H3K9me2 from transcribed genes (Figure 4F).

Although we thank the referee for this and other constructive suggestions, "apparent discrepancies" mentioned here are based on misunderstanding, we are afraid. The strongest point raised by this referee here is "in this manuscript the majority of the claimed direct targets of IBM1 are transposon genes", but we have not claimed that. What we claimed is opposite. Results in this manuscript, as well as those in our previous manuscripts, indicate that transcribed genes, rather than transposon genes, are targets of IBM1. The *ibm1* mutation affects DNA methylation of transcribed genes (Miura et al 2009) and H3Kme2 of transcribed genes (Inagaki et al 2010), rather than TE genes. In this manuscript, we show that the *ibm1* mutation mainly affects transcription of cellular genes, rather than TE genes (please see our response to point #1 of referee #1). It is true that ectopic H3K9me2 induces transcriptional repression in only a subset of genes (described here as DGs, which do not include the BONSAI gene), but effects on TEs are much less than that on genes. The results we describe in this manuscript and those in previous papers are fully consistent.

2), Although the author showed that the H3K4me1 levels decreased in a subset of genes in ibm1ldl2 double mutant compared to the imb1 single mutant, it was disappointing that no attempt was made to address the biological function of LDL2. More specifically is the LDL2 a bona fide H3K4me1 demethylase? Whether the K436A mutation indeed compromised the presumed LDL2 demethylase activity is questionable. There could be trivial explanations; the mutant phenotype could be caused by a change in protein stability and/or cellular localization. At least three experiments should be conducted to address these issues, a) In vitro/in vivo assay to detect the demethylase activity of LDL2, b) ChIP-Seq. to compare H3K4me1 levels between ldl2 mutant and WT. c) Western blot to analyse possible changes in the LDL2 (WT) and LDL2(K436A) protein levels in transgenic plants (Fig. S1E).

Our response: We agree with the referee that biological function of LDL2 in wild type background should be examined. As we described in the response to the point #9 of referee #1, we added ChiP-seq and RNA-seq results of the *ldl2* single mutant and control WT plants. Some TEs are affected in both H3K4me1 and transcription, supporting the model that targets of LDL2 are sequences with H3K9me2. Some genes are affected when they acquire H3K9me2 in *ibm1* background, but they are not affected in IBM1 wild type background, most likely because they do not have H3K9me2. We also tried LDL2 demethylase assay as we did before on IBM1 (Inagaki et al 2010 EMBO J.), but we have not been able to detect the demethylase activity in that system so far. We did not include that in the manuscript because the results do not lead us to any solid conclusion; it is very possible that additional host factor(s) are necessary for reconstruction of the demethylation system.

Minor concerns,

1), The use of abbreviations e.g. UGs/UTGs and DGs/DTGs has to be consistent between figure legend and labelling in the figure (cf. Fig 1G).

Our response: We corrected the inconsistency. We thank the referee for pointing that out.

2), The INTRODUCTION is too short and should be expanded. The authors have published several papers on IBM1 and not all readers are familiar with their considerable historical data. The authors should provide more background in the INTRODUCTION section.

Our response: We thank the referee that he/she appreciates our previous works on IBM1. We added explanation for properties of the *ibm1*-induced ectopic DNA methylation and H3K9me2 more thoroughly in the INTRODUCTION.

3), All data shown in the manuscript were not tested for statistical significance. This should be rectified.

Our response: We added statistical analyses for the results in new Figs 2C, 2D, S3 (described in the main text), S6B, S7A, and S8A.

Referee #3:

This referee is very positive. Nonetheless, the comments are very constructive and helped us to improve the manuscript very much.

Histone H3 lysine 9 methylation (H3K9me) is a repressive epigenetic mark that is found not only at promoter regions of transcriptionally silent genes but also on internal regions (gene bodies), where its biological significance is not well understood. As the authors have previously shown, mutations in IBM1, a histone H3 lysine 9 (H3K9) demethylase, result in ectopic H3K9me2 methylation at gene bodies and severe developmental defects in Arabidopsis. In the present study, they carried out a forward genetic screen on the ibm1 mutant to identify mutations that suppress the developmental defects in ibm1. This screen identified LDL2, a putative H3K4 demethylase as a mediator of the ibm1-induced developmental defects.
Further work using mRNA-seq and ChIP-seq to examine transcription and histone modifications on a genome-wide level in various mutant combinations led the authors to propose a model for differentiation of bi-stable epigenetic states at gene bodies through the activities of two histone demethylases: transcription induces loss of H3K9me2 through the action of IBM1, and H3K9me2 induces loss of H3K4me through the action of LDL2.

This study provides new and interesting information on how dynamic changes in histone modifications in gene bodies influence transcription. There are four LDL genes in the Arabidopsis genome and currently only functional information about one (LDL4 or FLD). This study adds new information on this important, conserved group of histone methyltransferases and will be interesting for epigeneticists working not only on plants but other systems as well.

1. Mention the name of the genes responsible for Type I mutations (CMT3 and SUVH4) in Results, not just in Materials and Methods.

Our response: As suggested, we added the name of genes for Type I mutations (CMT3, SUVH4, and HOG1) in Results.

2. The paper is quite short (only 3 figures in the main text). I recommend moving some important supplementary information into the main text. For example, show in the main text Venn diagrams currently in Fig S2, S3, Fig, 5A histone H3 methylation patterns in regions around DGs.

Our response: As suggested, we moved substantial parts of the results in Supplemental Figures to main Figures.

3. The experiments with suvh456 triple mutants, which are defective in three H3K9 methyltransferases that act redundantly to silence transposable elements, seem tacked on (e.g they are not mentioned in the Abstract). More effort could be made to integrate these results into the whole 'story' by adding something to the Abstract and Perspective section. Some results from the suvh456 experiments - e.g. Fig. S7 part A and Fig. S8 part A - could be shown in the main text.

Our response: As suggested, we added information related to those results to Abstract. In Abstract of the previous version, we described suvh456 results as “Furthermore, mutations of H3K9 methylases induced drastic increases in H3K4me1 in the bodies of diverse transposable elements”. To this sentence, we added “, and H3K4me1 increase appears to be a prerequisite for transcriptional de-repression of these TEs” in the revised manuscript. The added part corresponds to the new analysis shown in new Fig 6E. In addition, results on suvh456 in supplemental Fig S7 and S8 are moved to main Figures, as suggested.

3. The acronyms - DGs, UGs, DEGs, TEGs - get confusing. Except for DEGs, write out 'down-regulated' and 'up-regulated'. Refer to TEGs as 'TE genes'.

Our response: As suggested, we refer to TEGs as ‘TE genes’. We also corrected the inconsistency of DGs/DTGs and UGs/UTGs.

4. Although the authors aim 'To understand the molecular basis of the LDL2-mediated developmental defects' (p. 5), aside from identifying numbers of differentially expressed genes by mRNA-seq, there is no further insight into the developmental aspects purportedly being addressed in the study. Why are developmental defects of imb1 mutant suppressed in imb1 ldl2 if ectopic H3K9me is not removed in this double mutant? What is basis of development defects in ibm1? Further discussion of these points is needed; perhaps a GO analysis of the DEGs would help to shed some light on these questions.

Our response: We thank the referee for the very constructive comment. We did GO analysis of DGs as well as UGs. Very interestingly, the analyses revealed that pathogen response genes are highly enriched in UGs. We added this and other discussion related to the pathway to the developmental abnormalities in the first paragraph of the Perspective part (p.11).

2nd Editorial Decision 01 November 2016

Thank you for submitting a revised version of your manuscript. It has now been seen by two of the original referees and their comments are shown below.

As you will see the referees both appreciate the data and clarifications that you have provided to address the original concerns, but where ref #3 is now fully supportive of publication ref #1 lists a number of remaining criticisms - mostly linked to manuscript structure and data presentation - that you will have to address in a final revised version. In particular, you will see that the referee finds that the manuscript remains very densely written and that figures are presented out of order in several cases. Further to that, we had to notice that the manuscript is still very short (although the number of figures has increased) and that a more extensive presentation and discussion of the findings would make the data and conclusions more accessible for the readers. Since some of these points were already raised in the first round of review I would strongly encourage you to take them to heart at this point and extensively rewrite the manuscript. In addition to the text/presentation changes, referee #1 also requests that the statistical significance for the findings presented in figure 2D and 3 should be tested and discussed in the manuscript.

Based on the comments from the referees, I would like to invite you to submit a revised version of your manuscript in which you address the above concerns.

REFEREE REPORTS

Referee #1:

This is a review of the revised manuscript entitled "The gene-body chromatin modification dynamics mediates epigenome differentiation in Arabidopsis" submitted to the EMBO Journal by Inagaki et al. In this revision the authors respond to the points raised in the original version, and the data quality is improved compared to the original version of the manuscript. However, they have left this second version of the manuscript highly disorganized and very difficult to read.

For example, the figures are referred to in an out-of-order fashion throughout the manuscript. On page 6, Figure panels 2A-B, E, F-G are discussed, but 2C-D isn't discussed for a full page later and in a different context. This creates problems for the reader, as the terms UG and DG are used in the figures before they are ever defined in the text.

The main paragraph on page 8 is very difficult to understand / read. Every sentence refers to a figure at rapid-fire speed, and there are large skips or breaks in logic. In addition, the references are mostly to different figures. This is a critical paragraph of the manuscript and I suggest ordering the data in a linear and logical manner, rather than skipping between multiple figures and points without any explanation of logic.

Several paragraphs are very difficult to read because of disconnected sentences. For example, on page 6 the middle paragraph starts and ends discussing LDL2 function, but in the middle there is a sentence "Transcribed genes are affected by ibm1 mutation, and the IBM1 protein localizes to these transcribed genes, but not to silent genes or TEs." This sentence is then used as evidence that LDL2 mediates developmental phenotypes induced by gene-body heterochromatin in the ibm1 mutant. I think that this is very circumspect evidence for how the developmental phenotype is suppressed in LDL2.

Figure 2D is a weakness of the manuscript. It shows virtually no overlap between genes that have hyper H3K9me2 in ibm1, and the UG or DG genes. The overlap that is shown is likely not more than what is expected by chance (and this statistical test should be done). Therefore, this means that the UG and DG genes are indirectly up or down regulated, and are not direct targets of IBM1. But for the rest of the manuscript they are considered direct. The statistical test should be performed to determine if these genes are just random overlap of the datasets, and if so, they should not be considered and treated as direct targets.

Figure 3 is a weakness of the manuscript. I see only a slight difference in the ibm1/ldl2 double mutant compared to the ibm1 single mutant. Please perform a statistical analysis the shows that the difference discussed in the manuscript text are real.

As a general note, I suggest not using axes in scatter plots that are based in "change" such as in Figure 6. The authors plot change between two datasets vs. change in two other datasets. This is very difficult to parse what the authors are trying to show with the data, and the display should be simplified to make the large dataset understandable.

Referee #3:

The authors have addressed my comments in a satisfactory manner. I have no further suggestions for improvement.

2nd Revision - authors' response 14 November 2016

Point by point response to the comments by Referee #1

We are glad to see that the referee #1 regards that “In this revision the authors respond to the points raised in the original version, and the data quality is improved compared to the original version of the manuscript.” We also thank the referee for specific comments for further improving the manuscript. We have incorporated most of the comments.

On page 6, Figure panels 2A-B, E, F-G are discussed, but 2C-D isn't discussed for a full page later and in a different context.

We added the discussion about Fig 2C and 2D. Thank you for the suggestion.

The main paragraph on page 8 is very difficult to understand / read. Every sentence refers to a figure at rapid-fire speed, and there are large skips or breaks in logic. In addition, the references are mostly to different figures. This is a critical paragraph of the manuscript and I suggest ordering the data in a linear and logical manner, rather than skipping between multiple figures and points without any explanation of logic.

We agree that this is an important paragraph. We have separated this paragraph into two paragraphs and added more explanation, so that readers can follow this part more easily. We have also changed the order of explanations.

For example, on page 6 the middle paragraph starts and ends discussing LDL2 function, but in the middle there is a sentence "Transcribed genes are affected by ibm1 mutation, and the IBM1 protein localizes to these transcribed genes, but not to silent genes or TEs." This sentence is then used as evidence that LDL2 mediates developmental phenotypes induced by gene-body heterochromatin in the ibm1 mutant. I think that this is very circumspect evidence for how the developmental phenotype is suppressed in LDL2.

We agree that the description about IBM1 protein localization in the middle of this paragraph have made the logic unclear. We have removed the description about IBM1 protein localization from this part and described this observation only in the first paragraph of Discussion.

Figure 2D is a weakness of the manuscript. It shows virtually no overlap between genes that have hyper H3K9me2 in ibm1, and the UG or DG genes. The overlap that is shown is likely not more than what is expected by chance (and this statistical test should be done).

The statistical test have been done and described in legend of Figure 2D in the previous version of the manuscript. The differences were highly significant ($P<10^{-24}$).

Figure 3 is a weakness of the manuscript. I see only a slight difference in the ibm1/ldl2 double mutant compared to the ibm1 single mutant. Please perform a statistical analysis the shows that the difference discussed in the manuscript text are real.

We added a new panel in Figure 3 (panel C), to show that a large number of genes lost H3K4me1 in ibm1 and that the affected genes were much enriched in genes accumulating H3K9me2 in ibm1. We also show in that panel that the number of genes with H3K4me1 loss was much reduced in the ldl2 ibm1 double mutant. We have added the statistical test in the legend of Figure 3.

I suggest not using axes in scatter plots that are based in "change" such as in Figure 6.

Although we agree it desirable to use the data as close to the original as possible, we are showing the changes in Figure 6, because the scatter plot can only show relationship of two parameters.

3rd Editorial Decision 07 December 2016

Thank you for submitting the revised version of your manuscript, which has now been seen once more by ref #1 whose comments are included below.

As you will see, the referee is now overall satisfied with the statistical analysis conducted but still has a few remaining suggestions that would improve data presentation for the reader. I would encourage you to incorporate the suggestions made by the referee in a final version of the manuscript and I would like to emphasize that this would in our view strengthen the paper.

REFEREE REPORTS

Referee #1:

The manuscript by Inagaki et al has improved from the previous version. However, two of my key points need further work:

1. The written presentation of the data in the manuscript is generally good, with the exception of the entire page 8. This page is simply not readable. Almost every single sentence on page 8 refers to primary data, and are no connecting sentences that provide the reader with an understanding of the logic behind why the analysis was done. Page 8 reads like a bullet list of facts. I suggest adding a sentence or two between each figure reference to explain to the reader the logic of why an analysis is being done, and what the result means.

2. The presentation of Venn diagrams is still very difficult for the reader to understand. In 3C, I suggest making two Venn diagrams. The point that the authors are trying to make has nothing to do with the 123 genes in the center of the Venn diagram, but rather should be two Venn diagrams: one showing the overlap between Hyper H3K9me2 and decreased H3K4me1 in ibm1, and the second Venn diagram being Hyper H3K9me2 and decreased H3K4me1 in ibm1/ldl2. Then the reader can compare and see that one is more than the other. The authors never even mention the 3-way overlap of 123.

3. In addition, in figure 5 and 6 the authors use inconsistent Y-axis scales, which makes their data seem to have a larger effect. For example, in 6F once all the Y-axes are set to 50, I think the H3K4me1, me2 and me3 may look very similar.

Other improvements to the writing:
1. In the abstract: "Here we show that H3K9me-directed removal of H3K4..." should read "Here we show that H3K9me-associated removal of H3K4..."
2. In the abstract: "The ldl2 mutation suppressed the developmental defects..." would be improved by writing "The ldl2 mutation suppressed the expression and developmental defects..."
3. In the first paragraph of the introduction, it is unclear whether the claims made about H3K9me are specific to H3K9me2 or all H3K9me states.

3rd Revision - authors' response 14 December 2016

Point-by-point response to the comments by Referee 1

1. The written presentation of the data in the manuscript is generally good, with the exception of the entire page 8. This page is simply not readable. Almost every single sentence on page 8 refers to primary data, and are no connecting sentences that provide the reader with an understanding of the logic behind why the analysis was done. Page 8 reads like a bullet list of facts. I suggest adding a sentence or two between each figure reference to explain to the reader the logic of why an analysis is being done, and what the result means.

In the top paragraph of page 8, we have added explanation about the results for the effect of *ibm1* single mutation to H3K4me1, before moving to the results of the *ibm1 ldl2* mutant. We separated the second paragraph into three paragraphs, each describing explanation for the amount of the modifications in DGs, H3K9me2/3 in DGs, and conclusion, so that the story becomes easier to follow. We also added more explanation in each of them. In addition, we put different panel names to different graphs in Figure 4 and 6, so that the results to see can be pinpointed easily. We thank the referee for the suggestion.

2. The presentation of Venn diagrams is still very difficult for the reader to understand. In 3C, I suggest making two Venn diagrams. The point that the authors are trying to make has nothing to do with the 123 genes in the center of the Venn diagram, but rather should be two Venn diagrams: one showing the overlap between Hyper H3K9me2 and decreased H3K4me1 in ibm1, and the second Venn diagram being Hyper H3K9me2 and decreased H3K4me1 in ibm1/ldl2. Then the reader can compare and see that one is more than the other. The authors never even mention the 3-way overlap of 123.

The Venn diagram of the three circles gives more information than two separate diagrams of two circles, with less space. Most readers can easily deduce the latter from the former, we believe. The two Venn diagrams suggested by the referee do not show that most of the genes with decrease in H3K4me1 in *ibm1 ldl2* are included in the genes with decrease in H3K4me1 in *ibm1* single mutant. Although that is what expected, we would like to show that information to the readers, because that may be related to the nature of genes with decrease in H3K4me1 in *ibm1 ldl2*. We changed expression in the main text to incorporate this point. We thank the referee for the suggestion.

3. In addition, in figure 5 and 6 the authors use inconsistent Y-axis scales, which makes their data seem to have a larger effect. For example, in 6F once all the Y-axes are set to 50, I think the H3K4me1, me2 and me3 may look very similar.

The most important message in Figure 6FG is the difference of r in the three modifications, which is reflected in the shape. When comparing different modifications, shape (or relative value) is more important than the absolute value, because the absolute value can change depending on the antibody used. We selected those scales depending on the dynamic ranges of the datasets, to make the shapes easier to see. We therefore would like to keep the figures as they are.

Other improvements to the writing:
1. In the abstract: "Here we show that H3K9me-directed removal of H3K4..." should read "Here we show that H3K9me-associated removal of H3K4..."

We changed the expression as suggested.

2. In the abstract: "The ldl2 mutation suppressed the developmental defects..." would be improved by writing "The ldl2 mutation suppressed the expression and developmental defects..."

This part is just after description of suppressor screening, which is done by observing developmental defects and silent mark. We therefore would like to refrain from adding “the expression and” in this part, because that might be confusing.

3. In the first paragraph of the introduction, it is unclear whether the claims made about H3K9me are specific to H3K9me2 or all H3K9me states.

The first part of the abstract and introduction is about H3K9me in general. In mammals, Drosophila, and fission yeast, H3K9me3/2 are both silent marks. In plants, H3K9me2 is very well characterized, but not much evidence is reported for the function of H3K9me3. We therefore uses H3K9me for describing eukaryotic silent mark in general, and use H3K9me2 for specifically describing effects of Arabidopsis mutations such as *ibm1* and *suvh456*. We do not want to describe the details of H3K9me3 in this part, because it is still controversial, and it is irrelevant to the contents of this paper.

4th Editorial Decision 16 December 2016

Thank you for submitting the final revision of your study, I am pleased to inform you that your manuscript has now been accepted for publication in The EMBO Journal.

EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Tetsuji Kakutani
Journal Submitted to: The EMBO Journal
Manuscript Number: EMBO J-2016-94983

Reporting Checklist For Life Sciences Articles (Rev. July 2015)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

- the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
- graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- if n< 5, the individual data points from each experiment should be plotted and any statistical test employed should be justified
- Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.

- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple χ2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P values = x but not P values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

Please ensure that the answers to the following questions are reported in the manuscript itself. We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

In the pink boxes below, provide the page number(s) of the manuscript draft or figure legend(s) where the information can be located. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable).

B- Statistics and general methods

Please fill out these boxes ↓ (Do not worry if you cannot see all your text once you press return)

1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	Epigenomic analyses such as BS-seq and ChIP-seq are conducted for two biological replicates, and transcriptomic analyses are conducted for three biological replicates. We decided the sample sizes based on the field's standard.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	NA
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	NA
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	NA
For animal studies, include a statement about randomization even if no randomization was used.	NA
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	No.
4.b. For animal studies, include a statement about blinding even if no blinding was done	NA
5. For every figure, are statistical tests justified as appropriate?	All the statistical tests are justified as appropriate.
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	Kolmogorov-Smirnov test has no need to make assumption about the distribution of data (non-parametric). Hypergeometric test does not need any assumption about the distribution of values.
Is there an estimate of variation within each group of data?	Yes.
Is the variance similar between the groups that are being statistically compared?	Kolmogorov-Smirnov test tests if two distributions themselves are same or not, therefore variances do not have to be similar.

C- Reagents

USEFUL LINKS FOR COMPLETING THIS FORM

http://www.antibodypedia.com
http://1degreebio.org
http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repo

http://grants.nih.gov/grants/olaw/olaw.htm
http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm
http://ClinicalTrials.gov
http://www.consort-statement.org
http://www.consort-statement.org/checklists/view/32-consort/66-title

http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tun

http://datadryad.org

http://figshare.com

http://www.ncbi.nlm.nih.gov/gap

http://www.ebi.ac.uk/ega

http://biomodels.net/

http://biomodels.net/miriam/
http://jjj.biochem.sun.ac.za
http://oba.od.nih.gov/biosecurity/biosecurity_documents.html
http://www.selectagents.gov/

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	All the information (eg. catalog number and reference) about the antibody used are provided in the Materials and Methods section.
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	NA

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	NA
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	NA
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	NA

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	NA
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	NA
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	NA
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	NA
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	NA
16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	NA
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	NA

F- Data Accessibility

18. Provide accession codes for deposited data. See author guidelines, under 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	Sequence data are deposited to DNA Data Bank of Japan (DDBJ).
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	NA
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	NA
21. As far as possible, primary and referenced data should be formally cited in a Data Availability section. Please state whether you have included this section. Examples: **Primary Data** Wetmore KM, Deutschbauer AM, Price MN, Arkin AP (2012). Comparison of gene expression and mutant fitness in Shewanella oneidensis MR-1. Gene Expression Omnibus GSE39462 **Referenced Data** Huang J, Brown AF, Lei M (2012). Crystal structure of the TRBD domain of TERT and the CR4/5 of TR. Protein Data Bank 4O26 AP-MS analysis of human histone deacetylase interactions in CEM-T cells (2013). PRIDE PXD000208	NA
22. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	NA

G- Dual use research of concern

23. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines, provide a statement only if it could.	NA