THE EMBO JOURNAL

Manuscript EMBO-2016-95335

# PRC2 is dispensable for HOTAIR-mediated transcriptional repression

Manuela Portoso, Roberta Ragazzini, Živa Brenčič, Arianna Moiani, Audrey Michaud, Ivaylo Vassilev, Michel Wassef, Nicolas Servant, Bruno Sargueil and Raphaël Margueron

*Corresponding author: Raphaël Margueron, Institut Curie, INSERM U934, CNRS UMR3215*

**Review timeline:**

| | |
|---|---|
| Submission date: | 25 July 2016 |
| Editorial Decision: | 06 September 2016 |
| Revision received: | 18 December 2016 |
| Editorial Decision: | 21 December 2016 |
| Revision received: | 23 December 2016 |
| Accepted: | 05 January 2017 |

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 September 2016

Thank you for submitting your manuscript for consideration by the EMBO Journal and my apologies for the extended duration of the review time. Your study has now been seen by three referees whose comments are shown below.

As you will see from the reports, all three referees express great interest in the findings reported in your manuscript and highlight the importance of this data for the field. However, while ref #1 is fully supportive of the manuscript as is (pending minor textual clarifications) refs #2 and #3 raise a couple of concerns that will require some additional experimental work to address.

Given the referees' overall positive recommendations, I would 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 of your manuscript will therefore depend on the completeness of your responses in this revised version.

For the revised manuscript I would particularly ask you to focus your efforts on the following points:

-> Please include the textual clarifications suggested by ref #1

-> Please also include the additional ChIP-experiments requested by ref #2 to strengthen the conclusions in figs 3 and 5. I realize that this will involve additional work on your side but the referee is overall supportive of your work and wants to ensure that the conclusions cannot be questioned afterwards.

-> You will see that ref #3 would have liked to see more data on the actual mechanism by which HOTAIR silences target genes, but I also appreciate that this would be out of the scope of the

current study. However, I would ask you to address/clarify if DNA methylation or H3K9 methylation on the reporter would contribute to the effects seen here. In addition, I would ask you to comment/clarify the localization of the ChIP primers and discuss the possible structural effects of adding MS2 sites to HOTAIR.

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

------------------------------------------------

REFEREE REPORTS

Referee #1:

Many lncRNAs have been claimed to interact with PRC2 leading to a proposal that lncRNAs act to recruit PRC2 to its target sites on genomic DNA. While this model was exciting, the evidence has been weak and many recent papers have begun to cast serious doubts on its accuracy. In one of the most rigorous papers to date, Portosa et al. specifically dissects the HOTAIR lncRNA, one of the very first lncRNAs claimed to guide PRC2, and demonstrates that PRC2 is dispensable for HOTAIR function. It is difficult to overstate the importance of this paper, although it may appear to be a negative result, the implications in the lncRNA and chromatin fields will be enormous. Overall, I find the experiments performed to be well designed and controlled and the overall design extremely elegant. I particularly like the development of the exogenous recruitment assay and feel this can and should be a new standard for studying proposed lncRNA functions. Because of the important implications of this paper, I hope that it will be published as soon as is reasonably possible.

I have several minor comments that I hope the authors can address before publication in order to make it clearer and more accurate:

1) In the Abstract, the authors say that there is no "clear-cut evidence demonstrating this novel mode of regulation". I think this isn't entirely accurate since Xist is a clear case where the lncRNA modulates gene expression by recruiting chromatin proteins (just not PRC2). Perhaps it would be more accurate to say there is no clear-cut evidence that lncRNA-mediated recruitment of PRC2 leads to gene expression changes.

2) On page 3, the authors say that "a majority of lncRNAs is reported to influence transcription in the nucleus...". This is not accurate, the majority of lncRNAs are not functionally characterized at all. I would restate this as "several lncRNAs are reported to influence transcription".

3) On page 3, the authors state that "the variety of lncRNAs, their relatively low levels of expression and their tissue-specific pattern of expression all point toward potential functions in development". This argument does not make sense to me - how does low expression point to a role in development?

4) On page 3, the authors claim that recent studies aimed at identifying Xist interacting proteins did not find PRC2 components, but Minajigi et al. 2015 did actually find PRC2 components. But, it is still not clear that this is directly interacting. You should reword this.

5) In addition to the arguments for why PRC2 is likely indirectly recruited by Xist, I think it would be worth noting that deletion of PRC2 is not required for Xist-mediated silencing. This result is most consistent with the observations in the current paper.

6) In several places in the Introduction, Results, and Discussion, the authors claim that they show that HOTAIR can repress transcription in cis, but this not true. It seems that they are using the terms cis and trans incorrectly. When they say cis they mean when HOTAIR is recruited to an exogenous location, which is distinct from cis/trans. Specifically, cis refers to regulation on the same chromosome from which it is transcribed. More accurate terminology is needed here.

7) Can the authors mention how many genes changed when HOTAIR is overexpressed? They mention the criteria, but never the number. Also, 2-fold and $p<0.05$ isn't a particularly rigorous

cutoff. I understand why they chose this, but could they add 1 sentence including the numbers with a more reasonable FDR cutoff for the reader?

8) In Figure 2, Figure 3, , I struggled to interpret the bargraph legends (1,2,3,4) for several minutes. Could you please write a description - or at a minimum describe them in the legend - rather than rely on people connecting up the numbers across panels?

9) Figure 4C labels are pixelated and hard to read

Referee #2:

In this manuscript entitled "PRC2 is dispensable for HOTAIR-mediated transcriptional repression" the authors focus on a specific lncRNA called HOTAIR and its association with the polycomb repressive complex 2 (PRC2). HOTAIR has previously been linked with PRC2 recruitment to chromatin and reported to be overexpressed in breast cancers and potentially capable of driving the tumourigenisis by reprogramming PRC2 targeting in genome (Gupta et al, 2010).

Margueron and colleagues develop a novel RNA-chromatin tethering system in breast cancer cells and report that although HOTAIR overexpression induces H3K27me3 deposition and transcriptional repression at an exogenous transgene, this effect is intriguingly not mediated by PRC2. Although this manuscript has begun to address a very important issue in the Polycomb field and puts forward strong in vitro evidence, given the fact that the data refutes an already published dataset (Gupta et al 2010), the manuscript requires further strengthening and experimental evidence to further support their claims.

Major Comments;

1. In Figure 3, given the observation that the H3K27me3 enrichments at the luciferase transgene are quite modest compared to the endogenous MYT1 gene, the authors should perform PRC2 (Ezh2/Suz12/EED) ChIPs in these settings. If their interesting claim that transcriptional repression by HOTAIR is independent of PRC2 holds true, PRC2 should not be enriched at these sites.
2. Following on from this, in order to further mechanistically explain how HOTAIR induces transcriptional repression, the authors could perform H3K27Ac and RNA PolII ChIPs at the luciferase transgene promoter.
3. JARID2 has previously been reported to contain an RNA binding domain and function in the recruitment of PRC2 to chromatin (Kaneko et al, 2014 and da Rocha et al, 2014). It is unclear from the main manuscript whether the recombinant PRC2 used in Figure 4 for the in vitro experiments contains JARID2? They could include JARID2 ChIPs along with the PRC2 ChIPs, as requested in Figure 3.
4. The observation in Figure 5 that repression of the luciferase transgene occurs in the absence of a functional PRC2 complex is very interesting. However, this finding requires further experimental evidence. For example, the authors should perform H3K27me3 and PRC2 ChIPs (EED/Suz12/Ezh2) at the luciferase transgenes (and MYT1 as a positive control) in both the EED and SUZ12 KO cell lines. It could be argued that there is still low levels of bulk H3K27me3 remaining in these cells from the western blot, and so the suggested experiment would rule out whether any residual H3K27me3 in the knockouts could be causing this repression. In any case, it's a necessary negative control.

Minor Comments;

1. The data in Figure 1, specifically panel D is quite confusing and it was unclear what the intentions of this panel are. Is it simply to show RNA-seq read count in cells ectopically expressing HOTAIR is low? To address this, the authors could very simply improve the labelling on the lower panels in Figure 1.
2. Given the fact that HOTAIR RNA in an endogenous setting is an antisense transcript, the labelling of transgenic cell lines in Figure 2 panel A was confusing at first glance. To make this more accessible to the reader, the authors could modify their labelling of cell lines #3 and #4; e.g. "Gal4-MS2BP MS2-loops-HOTAIR +ve/-ve" or "Gal4-MS2BP MS2-loops-HOTAIR control / knockdown".

3. The authors should also include cell line #4 in Figure 2E as an extra negative control.
4. The error bars in Figure 3 are technically not the best and require strengthening to increase the robustness of the result.

References:

1. Gupta RA., Shah N., Wang KC., et al (2010) Long no-coding RNA HOTAIR reprograms RNA state to promote cancer metastasis. Nature 464:1071-6
2. Kaneko S., Bonasio R., Saldana-Meyer R., et al (2014) Interactions between JARID2 and non-coding RNAs regulate PRC2 recruitment to chromatin. Mol Cell 53: 290-300
3. Da Rocha ST., Boeva V., Escamilla-Del-Arenal M., et al (2014) JARID2 in implicated in the initial Xist-induced targeting of PRC2 to the inactive X chromosome. Mol Cell 53:301-316.

Referee #3:

In this manuscript, Portoso and colleagues address the question of how a long non-coding RNA - HOTAIR - mediates changes in gene expression. The functions of lncRNAs is a very important and topical area and, despite a lot of studies on the popular HOTAIR lncRNA - many using crude over-expression or knockdown approaches, its biological function and mechanism of action remains unclear and disputed.

Here the authors, use an elegant combination of approaches to investigate the ability of HOTAIR to repress genes in trans and to determine any dependence on PRC2 (which published studies have implicated in HOTAIR mediated effects). They first show that HOTAIR overexpression in a breast cancer cell line has only very modest affects of gene expression and that this is independent of PRC2 (using matched PRC2 knockout cells). They then set up an MS2-based RNA tethering system to specifically recruit HOTAIR to a reporter transgene (tk-luciferase) with UAS/Gal4 binding sites and show a strong silencing effect on the transgene that is independent of PRC2 activity. This is conclusive as far as it goes, but it is disappointing that the study does not go on to identify the repression mechanism that is responsible for the strong reporter gene silencing induced by HOTAIR tethering.

Major points related to the findings and conclusions of this paper are:

1. How relevant are these data obtained at an artificial reporter transgene to the mechanism of action of HOTAIR at endogenous loci? One issue is that this transgene - as is often the case for transgenes in mammalian cells - may be particularly prone to gene silencing pathways that are dependent on DNA methylation and H3K9 methylation. This would preclude a major influence of polycomb. The authors should assay the reporter before and after HOTAIR tethering for DNA methylation status and for H3K9me2/3. There should be some discussion about the pros and cons of assaying lncRNA mechanisms using artificial reporter genes.

2. It appears that the primers used for ChIP are all located within the luciferase gene body, and so the authors are not assaying the chromatin state at the gene promoter - probably the most important site to look at. ChIP should be performed for the tk promoter.

3. HOTAIR has a highly folded secondary structure (Somarowthu et al., 2015, Molecular Cell). It is possible that the addition of MS2 sites to HOTAIR, and in addition the binding of MS2BP to those MS2 sites, alters the secondary structure of HOTAIR and hence its function and protein interactions. Therefore, correctly the authors examine the structure of their HOTAIR molecules by SHAPE analysis. However, the data presented in Fig. 4C appear to only show the first 530bp of the molecule i.e not the region where the MS2 sites are located. Where is the data to show that the structure of the rest of the RNA molecule is not perturbed? The authors need to discuss the possibility that MS2BP binding may perturb HOTAIR structure.

1st Revision - authors' response 18 December 2016

© European Molecular Biology Organization 

Manuscript: EMBOJ-2016-95335

By Portoso *et al*.

First of all, we would like to thank the reviewers for their time and constructive comments on our manuscript. We tried to address most of the reviewer comments. We hope that it improves the manuscript and that it is now suitable for publication in EMBO.

**Referee #1:**

Many lncRNAs have been claimed to interact with PRC2 leading to a proposal that lncRNAs act to recruit PRC2 to its target sites on genomic DNA. While this model was exciting, the evidence has been weak and many recent papers have begun to cast serious doubts on its accuracy. In one of the most rigorous papers to date, Portosa et al. specifically dissects the HOTAIR lncRNA, one of the very first lncRNAs claimed to guide PRC2, and demonstrates that PRC2 is dispensable for HOTAIR function. It is difficult to overstate the importance of this paper, although it may appear to be a negative result, the implications in the lncRNA and chromatin fields will be enormous. Overall, I find the experiments performed to be well designed and controlled and the overall design extremely elegant. I particularly like the development of the exogenous recruitment assay and feel this can and should be a new standard for studying proposed lncRNA functions. Because of the important implications of this paper, I hope that it will be published as soon as is reasonably possible.

I have several minor comments that I hope the authors can address before publication in order to make it clearer and more accurate:

1) In the Abstract, the authors say that there is no "clear-cut evidence demonstrating this novel mode of regulation". I think this isn't entirely accurate since Xist is a clear case where the lncRNA modulates gene expression by recruiting chromatin proteins (just not PRC2). Perhaps it would be more accurate to say there is no clear-cut evidence that lncRNA-mediated recruitment of PRC2 leads to gene expression changes.

We agree with the reviewer and, indeed, we were implicitly referring to PRC2. We modified the text to be more accurate.

2) On page 3, the authors say that "a majority of lncRNAs is reported to influence transcription in the nucleus...". This is not accurate, the majority of lncRNAs are not functionally characterized at all. I would restate this as "several lncRNAs are reported to influence transcription".

The reviewer is right; we have corrected the text accordingly.

3) On page 3, the authors state that "the variety of lncRNAs, their relatively low levels of expression and their tissue-specific pattern of expression all point toward potential functions in development". This argument does not make sense to me - how does low expression point to a role in development?

The hypothesis was based on the similarity with developmental genes, which tend to have lower expression levels than constitutive genes and the assumption that high level of expression could indicate a more structural function. Yet, we agree that this argument is debatable and we removed it.

4) On page 3, the authors claim that recent studies aimed at identifying Xist interacting proteins did not find PRC2 components, but Minajigi et al. 2015 did actually find PRC2 components. But, it is still not clear that this is directly interacting. You should reword this.

Actually, we wrote "Xist interactome did not retrieve any **specific** PRC2 components". We meant to indicate that, components that are an exclusive signature of PRC2 (EED, SUZ12, EZH1/2) were not recovered. Only components (RBBP5 and 7) that are present in several complexes and do not allow concluding on the presence of PRC2 were found.
We have reformulated the sentence to avoid ambiguity.

5) In addition to the arguments for why PRC2 is likely indirectly recruited by Xist, I think it would be worth noting that deletion of PRC2 is not required for Xist-mediated silencing. This result is most consistent with the observations in the current paper.

We have added one sentence to refer to this notion.

6) In several places in the Introduction, Results, and Discussion, the authors claim

that they show that HOTAIR can repress transcription in cis, but this not true. It seems that they are using the terms cis and trans incorrectly. When they say cis they mean when HOTAIR is recruited to an exogenous location, which is distinct from cis/trans. Specifically, cis refers to regulation on the same chromosome from which it is transcribed. More accurate terminology is needed here.

We agree that our terminology was a bit of a leap. We have modified the text accordingly.

7) Can the authors mention how many genes changed when HOTAIR is overexpressed? They mention the criteria, but never the number. Also, 2-fold and p<0.05 isn't a particularly rigorous cutoff. I understand why they chose this, but could they add 1 sentence including the numbers with a more reasonable FDR cutoff for the reader?

We added a table (Figure EV1C) showing the number of differentially expressed genes at different cutoff (p<0;05 and p<0,01, logFC>1 and logFC>2).

8) In Figure 2, Figure 3, I struggled to interpret the bargraph legends (1,2,3,4) for several minutes. Could you please write a description - or at a minimum describe them in the legend - rather than rely on people connecting up the numbers across panels?

We have modified the scheme to make it easier to understand on figure 2. We also added the correspondence between number and cell line to all figures.

9) Figure 4C labels are pixelated and hard to read

We are sorry for this issue with figure resolution, it probably appeared when we reduced the size of the figure for submission.

**Referee #2:**

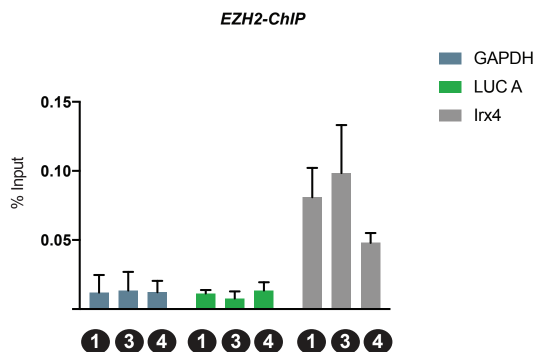
In this manuscript entitled "PRC2 is dispensable for HOTAIR-mediated transcriptional repression" the authors focus on a specific lncRNA called HOTAIR and its association with the polycomb repressive complex 2 (PRC2). HOTAIR has previously been linked with PRC2 recruitment to chromatin and reported to be overexpressed in breast cancers and potentially capable of driving the tumourigenisis by reprogramming PRC2 targeting in genome (Gupta et al, 2010).

Margueron and colleagues develop a novel RNA-chromatin tethering system in breast cancer cells and report that although HOTAIR overexpression induces H3K27me3 deposition and transcriptional repression at an exogenous transgene, this effect is intriguingly not mediated by PRC2. Although this manuscript has begun to address a very important issue in the Polycomb field and puts forward strong in vitro evidence, given the fact that the data refutes an already published dataset (Gupta et al 2010), the manuscript requires further strengthening and experimental evidence to further support their claims.

Major Comments;

1. In Figure 3, given the observation that the H3K27me3 enrichments at the luciferase transgene are quite modest compared to the endogenous MYT1 gene, the authors should perform PRC2 (Ezh2/Suz12/EED) ChIPs in these settings. If their interesting claim that transcriptional repression by HOTAIR is independent of PRC2 holds true, PRC2 should not be enriched at these sites.

We have done the requested experiment (see below) and indeed we cannot see EZH2 recruitment above the background. However, EZH2 ChIP is far less sensitive than H3K27me3, therefore we cannot conclude whether it is absent or below the detection limit.



Consequently, we prefer to not make any statement in the manuscript regarding this point.

2. Following on from this, in order to further mechanistically explain how HOTAIR induces transcriptional repression, the authors could perform H3K27Ac and RNA PolII ChIPs at the luciferase transgene promoter.

We have done the requested experiments, the ChIPs are now part of figure EV3b.

3. JARID2 has previously been reported to contain an RNA binding domain and function in the recruitment of PRC2 to chromatin (Kaneko et al, 2014 and da Rocha et al, 2014). It is unclear from the main manuscript whether the recombinant PRC2 used in Figure 4 for the in vitro experiments contains JARID2?

As indicated by the Coomassie gel staining of figure EV4, we used a 4 polypeptides PRC2 (EZH2, SUZ12, EED and RBBP4) to be consistent with previous reports.

They could include JARID2 ChIPs along with the PRC2 ChIPs, as requested in Figure 3.

Since we could not detect PRC2 at the reporter transgene, we assume that this request was not relevant anymore.

4. The observation in Figure 5 that repression of the luciferase transgene occurs in the absence of a functional PRC2 complex is very interesting. However, this finding requires further experimental evidence. For example, the authors should perform H3K27me3 and PRC2 ChIPs (EED/Suz12/Ezh2) at the luciferase transgenes (and MYT1 as a positive control) in both the EED and SUZ12 KO cell lines. It could be argued that there is still low levels of bulk H3K27me3 remaining in these cells from the western blot, and so the suggested experiment would rule out whether any residual H3K27me3 in the knockouts could be causing this repression. In any case, it's a necessary negative control.

We have done the requested control which is now included in figure 5.

Minor Comments;

1. The data in Figure 1, specifically panel D is quite confusing and it was unclear what the intentions of this panel are. Is it simply to show RNA-seq read count in cells ectopically expressing HOTAIR is low? To address this, the authors could very simply improve the labelling on the lower panels in Figure 1.

We modified the figure to better convey the message that genes, differentially expressed upon HOTAIR overexpression in MDA-MB-231, are characterized by very low read count.

2. Given the fact that HOTAIR RNA in an endogenous setting is an antisense transcript, the labelling of transgenic cell lines in Figure 2 panel A was confusing at first glance. To make this more accessible to the reader, the authors could modify their labelling of cell lines #3 and #4; e.g. "Gal4-MS2BP MS2-loops-HOTAIR +ve/-ve" or "Gal4-MS2BP MS2-loops-HOTAIR control / knockdown".

We changed the labeling of the transgenic cell lines to take this comment into consideration.

3. The authors should also include cell line #4 in Figure 2E as an extra negative control.

Due to time constraint, we could not address this question. However, we believe that we already showed the specificity of the shGAL4-mediated release of transcriptional repression with the two already included negative controls.

4. The error bars in Figure 3 are technically not the best and require strengthening to increase the robustness of the result.

We agree that the error bars are not very small but it reflects true biological replicates and the relatively low enrichment of H3K27me3 upon artificial tethering of MS2 HOTAIR RNA. In addition, due to an antibody whose commercialization was temporary interrupted, figure 3 and figure EV3 were performed with distinct antibodies and are therefore completely independent. Consequently, the ChIP for H3K27me3 was done more than 5 times. Therefore, we believe that this result is already really robust.

References:

1. Gupta RA., Shah N., Wang KC., et al (2010) Long no-coding RNA HOTAIR

reprograms RNA state to promote cancer metastasis. Nature 464:1071-6
2. Kaneko S., Bonasio R., Saldana-Meyer R., et al (2014) Interactions between JARID2 and non-coding RNAs regulate PRC2 recruitment to chromatin. Mol Cell 53: 290-300
3. Da Rocha ST., Boeva V., Escamilla-Del-Arenal M., et al (2014) JARID2 in implicated in the initial Xist-induced targeting of PRC2 to the inactive X chromosome. Mol Cell 53:301-316.

Referee #3:

In this manuscript, Portoso and colleagues address the question of how a long non-coding RNA - HOTAIR - mediates changes in gene expression. The functions of lncRNAs is a very important and topical area and, despite a lot of studies on the popular HOTAIR lncRNA - many using crude over-expression or knockdown approaches, its biological function and mechanism of action remains unclear and disputed.

Here the authors, use an elegant combination of approaches to investigate the ability of HOTAIR to repress genes in trans and to determine any dependence on PRC2 (which published studies have implicated in HOTAIR mediated effects). They first show that HOTAIR overexpression in a breast cancer cell line has only very modest affects of gene expression and that this is independent of PRC2 (using matched PRC2 knockout cells). They then set up an MS2-based RNA tethering system to specifically recruit HOTAIR to a reporter transgene (tk-luciferase) with UAS/Gal4 binding sites and show a strong silencing effect on the transgene that is independent of PRC2 activity. This is conclusive as far as it goes, but it is disappointing that the study does not go on to identify the repression mechanism that is responsible for the strong reporter gene silencing induced by HOTAIR tethering.

Major points related to the findings and conclusions of this paper are:

1. How relevant are these data obtained at an artificial reporter transgene to the mechanism of action of HOTAIR at endogenous loci? One issue is that this transgene - as is often the case for transgenes in mammalian cells - may be particularly prone to gene silencing pathways that are dependent on DNA methylation and H3K9 methylation. This would preclude a major influence of Polycomb. The authors should assay the reporter before and after HOTAIR tethering for DNA methylation status and for H3K9me2/3. There should be some discussion about the pros and cons of assaying lncRNA mechanisms using artificial reporter genes.

We agree that the tethering system we set up has limitations and we carefully referred to it as "artificial tethering" and "model cell lines" yet we believe that we have provided a variety of controls to insure that it is relevant. Also, it is one among several approaches (e.g. in vitro and HOTAIR overexpression) which all

point toward the same conclusion. Regarding DNA methylation, we have performed the requested experiments (Figure EV3c) and we could not detect any sign of DNA methylation at the transgene.

2. It appears that the primers used for ChIP are all located within the luciferase gene body, and so the authors are not assaying the chromatin state at the gene promoter - probably the most important site to look at. ChIP should be performed for the tk promoter.

We used different set of primers depending on what we aim to detect. The set of luciferase primers labeled "LUC A" is indeed located on the luciferase gene body but very close to the 5XUAS/tk promoter (less than 20nt away downstream of the tk promoter). Therefore, it reflects what happen at the promoter as illustrated by the fact that only this set of primers detects the GAL4-MS2BP (Figure 2b). In contrast, the enrichment for H3K27me3 upon tethering of HOTAIR with this set of primers is low (Figure 3a). A significant enrichment for H3K27me3 is only observed further downstream (set of primers LUC D and LUC E).

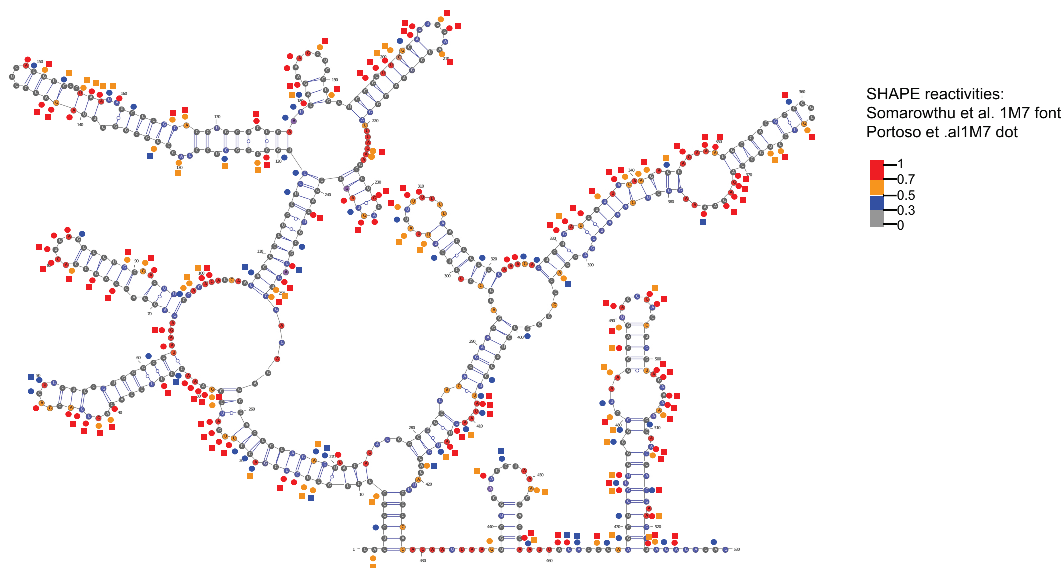
3. HOTAIR has a highly folded secondary structure (Somarowthu et al., 2015, Molecular Cell). It is possible that the addition of MS2 sites to HOTAIR, and in addition the binding of MS2BP to those MS2 sites, alters the secondary structure of HOTAIR and hence its function and protein interactions. Therefore, correctly the authors examine the structure of their HOTAIR molecules by SHAPE analysis. However, the data presented in Fig. 4c appear to only show the first 530bp of the molecule i.e not the region where the MS2 sites are located.

The SHAPE analysis was done to check that HOTAIR was correctly folded for the *in vitro* experiments; therefore the experiment was done with *in vitro* transcribed HOTAIR without any tag. Previous studies have reported that PRC2 binding to HOTAIR occur through its 5' region, we therefore focused on this region checking the independent structural module D1. Nonetheless, the full SHAPE reactivities are provided as sourced data for Figure 4c.

Where is the data to show that the structure of the rest of the RNA molecule is not perturbed?

We understand the referree's concern although the MS2 sites sequences fold into stable hairpin loops that should fold independently and are unlikely to interefere

with Hotair structure. However, to answer the referee's question, the HOTAIR sequence tagged with MS2 binding sites was submitted to the SHAPE-Map protocol using 1M7, and the reactivity map obtained was compared to the HOTAIR reactivity map previously obtained. The two reactivity maps are not identical but very similar, this was established in two different ways. First HOTAIR-MS2 reactivity map was used as constraint to model the structure of the 430 first nucleotides of HOTAIR that constitute the first structural domain that was supposed to bind PCBP. The most stable models obtained were mostly identical to those obtained using HOTAIR reactivity map, for which one example is shown in sup. Figure EV5. Second to assess the compatibility of HOTAIR-MS2 structure with the model established by Somarowthu et al., as we previously did for HOTAIR, we draw the reactivity on the Somarowthu's secondary structure model. This shows that reactivities of both HOTAIR and HOTAIR-MS2 are essentially compatible with each other and with Somarowthu's model. This strongly suggests that HOTAIR and HOTAIR MS2 domain I folds in the same way and that the MS2 repeats to do not alter HOTAIR its folding.




RNAstructure (Mathews lab) on HOTAIR Domain 1
Somarowthu et al. 2015

As exemplified in Somarowthu et al. modeling such a long RNA is a study in itself, which is way beyond the scope of this manuscript. Here we only use the SHAPE technology to show that HOTAIR domain one is correctly folded, and we did not model the rest of HOTAIR RNA. However, a quick comparison

of the reactivity map in between nucleotide 430 and 1580 suggests that HOTAIR and HOTAIR-MS2 fold in a similar way.

The authors need to discuss the possibility that MS2BP binding may perturb HOTAIR structure.

We added a sentence to clarify this point.

2nd Editorial Decision 21 December 2016

Thank you for submitting a revised version of your manuscript. It has now been seen by one of the original referees and this person's comments are shown below.

As you will see the referee finds that all criticisms have been sufficiently addressed; however, before we can go on to officially accept your manuscript for publication there are a few editorial issues concerning text and figures that I need you to address:

-> Please make sure that figures with error bars are only displayed when statistical testing is relevant (ie n equal to or larger than 3). For figures where data derives from n=2 we would ask you to show the two sets of data points instead (fig 1A, fig 2C and others) or to include data for an additional replica. This is needed for both main and Expanded view figures.

------------------------------------------------

REFEREE REPORTS

Referee #2:

I read the author's excellent rebuttal letter and I'm fully satisfied that they have addressed all the points I raised. I agree with Reviewer 1 that while this paper has "negative results", their implications in the lncRNA and chromatin fields are very significant. This paper should become well cited as a consequence.

3rd Editorial Decision 05 January 2017

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

© European Molecular Biology Organization 

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: Raphael Margueron |
|---|
| Journal Submitted to: EMBO Journal |
| Manuscript Number: EMBOJ-2016-95335 |

## 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).**

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/

### B- Statistics and general methods

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

| Question | Answer |
|---|---|
| 1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size? | All experiments were reproduced with completely independent biological replicates. Since this study involves the comparison of different cell lines, action were taken to avoid confounding factor such as growth condition. In particular, all experiments were done at different time with independent batch of cells. Sample size was increased when variance inter-experiment was observed. |
| 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? | Experiments were excluded when controls (positive or negative) did not folow the expected trend. |
| 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. | No randomization was performed. |
| 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. | Several key experiments were reproduced by independent experimentators. |
| 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? | yes |
| Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it. | The statistical analysis and the differential expression analysis of the RNA seq data was performed with the limma packages, as described in Material and Methods page 25 of the draft. |
| Is there an estimate of variation within each group of data? | Normalization was performed to take into account variation between each group of data (see Material and Methods for RNA-seq) |
| Is the variance similar between the groups that are being statistically compared? | Correlation between each replicates of each group is provided in figure EV1 |

### C- Reagents

| | |
|---|---|
| 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). | List of antibody with their origin is provided in appendix. |
| 7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination. | Origin of cell lines is provided, cells were tested for mycoplasma on montly basis. |

* 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'.<br><br>Data deposition in a public repository is mandatory for:<br>a. Protein, DNA and RNA sequences<br>b. Macromolecular structures<br>c. Crystallographic data for small molecules<br>d. Functional genomics data<br>e. Proteomics and molecular interactions | Accession code is provided: GSE72524 |
| 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.<br><br>Examples:<br>**Primary Data**<br>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<br>**Referenced Data**<br>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<br>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 |