An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment

Magnus Lynch, Andrew Smith, Marco De Gobbi, Maria Flenley, Jim Hughes, Douglas Vernimmen, Helena Ayyub, Jacqueline Sharpe, Jackie Sloane-Stanley, Linda Sutherland, Stephen Meek, Tom Burdon, Richard Gibbons, David Garrick, Douglas Higgs

Corresponding author: Douglas Higgs, John Radcliffe Hospital

Review timeline:
Submission date: 16 March 2011
Editorial Decision: 26 April 2011
Revision received: 15 August 2011
Editorial Decision: 20 September 2011
Revision received: 05 October 2011
Accepted: 13 October 2011

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 26 April 2011

Thank you for submitting your manuscript for consideration by The EMBO Journal, and please excuse the delay in getting back to you with the results of its external evaluation. I am herewith finally sending you the comments of three expert referees that had agreed to assess the study. All of them acknowledge the importance of the topic and the elegance and suitability of your experimental approach. Nevertheless, the reviews raise a number of major concerns that in our view preclude publication in The EMBO Journal at least in the present form. On close consideration of these points, I feel that we could potentially consider a revised manuscript further for publication if you should be able to adequately address at least certain key points (as well as of course all the specific technical and presentational issues raised). Among these key issues is the extension of the analysis in Figure 6 currently limited to only two loci (see ref 2 point 2, ref 3 point 7), ideally using an unbiased H3K27Me3 ChIP-Seq approach in conjunction with available datasets on polycomb target genes. Similarly, I would also recommend to heed referee 3's suggestion to extend some of the other site-specific analyses to a more significant number of locations, even though I feel that full genome-wide ChIP-seq analysis for Figures 1-4 as asked for by this referee may well be beyond the scope of the current study. Among the technical issues, it will be particularly important to improve the large-scale dataset analyses as requested by referee 3; and to properly normalize and control ChIP data (ref 2 minor point 1, ref 3 points 2 and especially point 3). On the other hand, I do not feel that the experiment asked for in referee 2's major point 1 would be strictly required to improve the message of the paper (while the point should still be diligently answered).
I realize that addressing these points may require substantial further efforts, and would like to remind you that it is our editorial policy to allow a single round of major revision only - eventual acceptance will thus depend on the completeness of your response and the conclusiveness of the added experimental work in the revised manuscript. If needed, we are happy to extend the revision period up to a total of six months, and I should emphasize that competing manuscripts appearing elsewhere during this time would not have a negative impact on our final assessment of your own study. However, we request that you contact the editor as soon as possible upon publication of any related work (to discuss how to proceed), or if you should foresee the need for extending the default three-month revision deadline.

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://www.nature.com/emboj/about/process.html

Please do not hesitate to contact me should you have any additional questions regarding this decision or your revision!

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

Yours sincerely,
Editor
The EMBO Journal

---

REFEREE REPORTS:

Referee #1 (Remarks to the Author):

Lynch et al
An interspecies analysis reveals determinants of the bivalent chromatin state

This manuscript addresses an important aspect of the functional determinants of chromatin states at genes. They use a human / mouse chromosome model they have previously established to demonstrate that specific DNA sequences located in cis are required for the establishment of bivalent marks. Furthermore their data indicates that the base composition of these sequences might be an important but not sufficient determinant of these states.

This is a significant study that addresses an important aspect of molecular biology - namely the importance of the underlying DNA sequences and possibly base compositions.

Specific comments
Fig 2B
It would be useful if the authors could show the EZH2, H3K27me3 and H3K4me3 ChIP data for the corresponding mouse Hba2 locus in these hybrid cells for comparison

Figure 3
How does the base composition of fragments (i) - (iv) compare?

Figure 4A
The authors state that "CpG density is generally conserved between corresponding regions in human and rodent genomes and this is associated with conservation of H3K27me3 recruitment in ES cells". To me this is not clear in this panel. Could the authors display correlation plots comparing human CpG density vs mouse CpG density and human H3K27me3 vs mouse H3K27me3 for their 2088 identified H3K27me3 peaks instead? The data points for the H3K27me3 plot could then be colour coded to distinguish between >3% CpG density and <=3% CpG density.
Referee #2 (Remarks to the Author):

In this paper Lynch et al. address an important unresolved question in developmental biology. That is, how are Polycomb group proteins recruited to their target DNA in mammalian cells? To do this they decided to study the alpha-globin gene (HBA) locus, which they observed was previously shown to be bound by Polycombs in human, but not mouse ES cells. From this they speculated that CpG islands are responsible for recruiting the Polycombs. To begin to test this, they inserted a 4kb sequence, encompassing the entire human HBA gene, into the orthologous position in the mouse locus and show by chromatin IP that it is capable of recruiting Polycomb group proteins. However, unfortunately they do not insert either the DNA fragments up or down stream of the CpG island within this 4kb fragment, to support their claim that the CpG island is required. They go on to mine previously published genome wide datasets of H3K27me3 enrichment profiles in human and mouse ES cells. In doing this, they nicely establish that the Polycomb target genes in human ES cells that are not enriched with H3K27me3 in mouse ES cells, contain CpG islands in human, but not in mouse cells. Finally, they also show, albeit on just two genes, that the loss of Dnmt1 or Dnmt3a/3b is correlated with a 2-fold increase in H3K27me3, but no change in H3K4me3.

Overall, the methods and approach of this paper is original, employing a powerful method of inserting DNA fragments into mouse ES cells in a controlled number of copies and performing a comparative study of chromatin states. However, the work relies heavily on data mining of previously published work and does not go far enough to warrant publication in EMBO J at this point.

Major points:

1. In Figure 3, the authors should test if the regions of DNA up stream and downstream of the CpG island are capable of recruiting Polycomb group proteins. If their hypothesis is correct, only the middle CpG island region should be capable of recruiting Polycombs.

2. Panels I-L of Figure 6 demonstrate that two genes, Ebf3 and Zfp423, have a two fold greater enrichment of H3K27me3 on their gene loci in Dnmt1 null and Dnmt3a/3b null cells. While this is not a very significant change, it could support their hypothesis that loss of methylation on CpG islands leads to an increased presence of Polycomb group proteins. It would significantly strengthen their argument if they were to perform a genome wide CHIP-SEQ analysis of H3K27me3 in these WT versus knock cells.

Minor points:

1) Several figures in this paper present CHIP analysis that incorrectly use "relative enrichment" as opposed to the accepted standard of "percentage input". These figures are Figure 1(C-F), Figure 2 (B, C, D), Figure 3 (B), Figure 4 (D, C, E), Figure 6 (I, J, K, L). These figures should be re-made to present the positive and negative controls, e.g. GATA6 and Actin.

2) Legend to Figure 1 (B): It is not clear from the legend how the region was replaced. Wallace et al, (2007) should be referenced, as in the main text, or the method explicitly stated.

3) Results section, page 7, paragraph two: no reference or evidence included for FERD3L being associated with bivalent chromatin.

4) Figure 3A: a scale indicating the fragment size would add to this panel.

5) Figure 5A: It would help the reader if the key genes in this paper were highlighted to the right on panel A, e.g. MYO1G, CLEC4G, MYF6 and GATA6.

6) Figure Legend 6(I-L): Two minor omissions from this figure legend. Firstly, the positive controls, Gata6 (for the H3K27me3 ChIP) and Beta Actin (for the H3K4me3 ChIP) are not mentioned. Secondly, the cell type should be stated in the legend.

7) The y-axis lines are missing in the ChIP-SEQ data in Figure Panels 3C, 5D and 6A.
8) Figure 6A, although it is mentioned in the main text that the intragenic RHBDL1 gene region is methylated in both human ES and somatic cells, this data should be presented in the Figure itself and also described in the Figure legend. It would also be appropriate to detail the cell type that the Cytosine methylation data was obtained from.

Referee #3 (Remarks to the Author):

"An interspecies analysis reveals determinants of the bivalent chromatin state" Lynch et al. investigates the bivalent chromatin state using a humanized mouse model. They show that differences in bivalency at the HBA2 and FERD3L loci between human and mouse ES cells are dominated by differences in regulatory sequences. Analysis of two non-overlapping regions of the HBA2 locus reveals that the bivalent state appears to be, at least in part, independently encoded in both fragments, suggesting direct recruitment of PcG proteins to multiple sites of a bivalent domain rather than spreading of said domain from an initial region of recruitment. Furthermore, the authors show that PRC2 occupancy is negatively correlated with transcriptional activity of HBA2. They computationally re-analyze public data to see whether CpG island erosion is associated with loss of H3K27me3 occupancy. The authors find that CpG islands that are marked with H3K4me3 or H3K27me3 are strongly biased towards unmethylated CpG islands and that demethylation of two analyzed CpG islands is accompanied with an increase in H3K27me3 but not H3K4me3 occupancy.

Given the important role of PcG proteins in developmental gene regulation and the limited knowledge of its regulation, the investigated topic is potentially of high interest. The recombination mediated cassette exchange approach, which controls location and copy number of inserted human DNA fragments into a mouse background, is elegant and well suited for the detailed analysis of gene regulation and chromatin state.

My single biggest criticism is that this elegant tool is not used to its fullest potential, because the authors consistently take single-locus approaches over much more persuasive global analyses. For instance, Figs 1-4 are in essence a series of bar-charts for about 4 site-specific ChIP reactions. It is difficult to extrapolate to the entire 120+ kilobase locus they are interrogating from these few data points. For an EMBO Journal calibre paper, I would hope to see the opposite: namely, ChIP-seq data in Figures 1-4, confirmed by site-specific analysis. Genome-wide data can also much more scientifically compelling and persuasive, because the transcriptional state of the entire human domain inserted into mouse would be presented within the figure, and the texture of the data would be clearer. Bar charts are rarely, if ever, intuitive.

Overall, most of the global conclusions are inferred from a limited number of experimental points, and thus their actual generality is difficult to evaluate. The paper would be tremendously strengthened if these elegant mouse constructs were interrogated using genome-wide tools, and the literature datasets were tested using genome-wide experiments.

Specific comments:

1. The authors state: "(H)owever the human a-globin locus is associated with prominent sites of PcG recruitment and chromatin bivalency in pluripotent cells whereas the corresponding mouse locus is not (Garrick et al, 2008)". Unfortunately, the previous publication by the authors (Garrick et al, 2008) does not provide this evidence. The only relevant comment I found in the 2008 paper is the claim in the discussion stating: "It is of interest that, unlike the human a-globin genes, we do not detect binding of the PRC2 complex and H3K27me3 modification at the orthologous mouse a-globin genes in nonexpressing cell types (data not shown)." When I realized that their prior publication was not about interspecies comparisons at all (in contrast to how they pitched Garrick 2008 in the intro), I felt like they needed to show in detail the human versus mouse differences before (or perhaps, within?) exploring it within this current Lynch et al manuscript.

2. The authors use ChIP-seq data from Ku et al. to support the difference in bivalency between the human and mouse HBA locus. The Ku data are presented in such a large window that it is difficult to evaluate in detail the chromatin state at the a-globin locus. It would be more convincing to also include data for other appropriate chromatin marks that are available from the same publication (H3K4me3, H3K27me3, etc.). Indeed, when I inspected the HBA locus in more detail using UCSC,
although the locus shows chromatin marks consistent with bivalency, the chromatin state of HBA1 and HBA2 can not be cleanly evaluated due to alignability issues. I would like to see a more explicit and detailed description of how the a-globin locuses align in mouse and human, maybe in supplemental. This issue needs to be carefully addressed when ChIP-seq data are used to evaluate occupancy patterns at repetitive genomic locations.

2. The ChIP-qPCR data in Figures 1 C-F and 4 B-E lack negative control regions, as is shown for example in Figure 2 B-D.

3. Throughout the manuscript the authors do not attempt to normalize the levels of H3K4me3 and H3K27me3 to bulk H3 occupancy. The observed changes in H3K4me3 and H3K27me3 occupancy could be explained by an increase in methylation levels, or alternatively by an increase in general nucleosome occupancy at a given location.

4. From the experiments presented in Figure 3 the authors conclude that the bivalent chromatin state is encoded multiple times within the HBA2 locus. This is an important point of the paper, and could be a major finding, but the experimental design has certain limitations:
   i) All fragments analysed overlap a specific CpG island. Thus the presented data would also be consistent with recruitment of PRC2 to the CpG island and spreading from there. In other words; would fragment (ii) show bivalency when significantly shorter and thus not overlapping the CpG island at the HBA2 promoter?
   ii) Fragment (ii) was analyzed at only a single point. This point showed marginal-to-no enrichment of Ezh2 and H3K27me3 in Figure 2B but the conclusions from this paragraph rest on one data point. This is related to the advantage of using ChIP-seq as primary data, versus confirmation data, above.
   iii) The ChIP-seq data in Figure 3 C are presented without an input control and lack experiments for H3K27me3 and H3K4me3. Also the alignability issues as mentioned in point 1 have to be addressed.

5. From the experiment in Figure 4 the authors conclude: "There is a linear, inverse relationship between H3K27me3 and H3K4me3 (Figure 4C-F) with the highest level of PRC2 recruitment and lowest level of H3K4me3 seen when the activating sequences associated with the promoter were deleted. Complete clearing of PRC2 and H3K27me3 occurred in the presence of the MC1 promoter and an intermediate level of H3K4me3 and H3K27me3 was observed for the wild type construct." To assume a linear relationship from only three data points is perhaps overly confident. By showing relative enrichment, rather than absolute enrichment relative to a genomic control known to be not enriched, it is not possible to evaluate if there is "complete clearing of PRC2 and H3K27me3". In general, all of the presented ChIP-qPCR results would be easier to evaluate and compare if they were presented as absolute rather than relative enrichment.

6. Their analysis of CpG data from other groups (mostly Figure 5) was not strong. Singling out <10% of the CpG islands that vary between human and mouse is not convincing, nor were the three, hand-selected examples where the CpG changes correspond to changes in the methylation status (Fig 5D-F).

7. The number of loci (2) analysed in Figure 6G to L is too small. These experiments need to be repeated with a larger sample.

8. How do the authors explain that there is no increase in H3K4me3 occupancy in Figures 6J and L since H3K27me3 alone occupies only a minority of CpG islands in ES cells (Ku et al Figure 4A).

9. The numbers indicating the scale of the displayed ChIP-seq data in Figure 6A are very small. Also data sets from the same factors in different cell types should be shown at the same scale and normalized for sequencing depth in order to be comparable.

Minor points:

1. On page 9 "Fragment 1" should be Fragment (i).
2. 'Salk Institute' or 'Broad Institute' are not appropriate references. The authors should provide references to the relevant papers.
3. Scales missing for heatmaps in Figure 5A.
4. All primer and probe sequences should be in a supplemental file.
5. In Figures 1 and 3 "Ku et al data" should be "Ku et al. data".

We would like to thank you for the opportunity to resubmit our revised manuscript to be considered for publication in the EMBO Journal. We are pleased to report that the experiments that you advised have substantially added to the strength of the paper.

As requested we have performed genome-wide chip-seq analysis for H3K27me3 in Dnmt3a/b -/- and wild type ES cells. Using bisulphite sequencing we have shown that CpG islands whose DNA is methylated in wild type ES cells lose methylation in Dnmt3a/b -/- ES cells. This loss of methylation is associated with recruitment of the PcG complex at these CpG islands in parallel with modification of the associated chromatin by H3K27me3. Interestingly many of these sites correspond to gene exons where the CpG density may reflect codon bias suggesting that a function of DNA methylation may be to prevent these sites from being recognised as CpG islands. This is an entirely novel finding which we are sure will be of considerable interest to the field and greatly strengthens our hypothesis that the density of unmethylated CpG dinucleotides is a key determinant of vertebrate PRC2 recruitment.

We feel that our revised manuscript addresses all of the referees’ comments and these are considered in detail below. With reference to the specific points that you highlighted in your letter, ChIP-qPCR data are now presented as percentage of input material and additional positive and negative control points have been added as requested. ChIP seq data are now normalized to mapped reads per 10 million to facilitate comparison between experiments. In response to referee 3’s request to extend the site specific analysis of RMCE chip data, we have added three qPCR points in the endogenous mouse alpha globin locus and extra points in the HBA2 and HBB fragments (Figures 2,3). We have also added unmodified H3 chip data to address referee 3’s concerns regarding histone occupancy (Figure S1G-I). However we agree that a full genome wide analysis for Figures 1-4 would be beyond the scope of this study since the introduced fragments are short (2-4kb in size) and are already densely covered by qPCR amplicons.

Response to Referee 1

This manuscript addresses an important aspect of the functional determinants of chromatin states at genes. They use a human / mouse chromosome model they have previously established to demonstrate that specific DNA sequences located in cis are required for the establishment of bivalent marks. Furthermore their data indicates that the base composition of these sequences might be an important but not sufficient determinant of these states.

This is a significant study that addresses an important aspect of molecular biology - namely the importance of the underlying DNA sequences and possibly base compositions.

Specific comments
Fig 2B
It would be useful if the authors could show the EZH2, H3K27me3 and H3K4me3 ChIP data for the corresponding mouse Hba2 locus in these hybrid cells for comparison

Data has been added for three additional qPCR probes situated in the mouse alpha globin locus including a probe specific to exon 1 of the mouse Hba2 and Hba1 genes. Please note that due to the very high degree of homology it is not possible to design qPCR probes that specifically quantify enrichment at Hba2 versus Hba1. Data for these new probes has been added into Figure 2 and Figure 3 as requested.

Figure 3
How does the base composition of fragments (i) - (iv) compare?

The base composition and CpG content of fragments (i)-(iv) along with the same data for the FERD3L and HBB fragments are now illustrated in Figure S2.

Figure 4A
The authors state that "CpG density is generally conserved between corresponding regions in human and rodent genomes and this is associated with conservation of H3K27me3 recruitment in ES cells". To me this is not clear in this panel. Could the authors display correlation plots comparing human CpG density vs mouse CpG density and human H3K27me3 vs mouse H3K27me3 for their 2088 identified H3K27me3 peaks instead? The data points for the H3K27me3 plot could then be colour coded to distinguish between >3% CpG density and <=3% CpG density.

As requested, this analysis has been performed and the data are displayed in Figure S3.

Figure 6
In the abstract the authors state "..chromatin bivalency reflects competition between Polycomb complex recruitment and transcriptional activation and that this is restricted to unmethylated CpG islands". For me it would help if the authors could include the equivalent panel to 6D showing H3K27me3 and H3K4me3 for unmethylated (<3% methylation) CpG islands similar to what is shown in Figure S2 but not discriminating for CpG island size.

These data are now shown in Figure S4A.

Response to Referee 2

In this paper Lynch et al. address an important unresolved question in developmental biology. That is, how are Polycomb group proteins recruited to their target DNA in mammalian cells? To do this they decided to study the alpha-globin gene (HBA) locus, which they observed was previously shown to be bound by Polycombs in human, but not mouse ES cells. From this they speculated that CpG islands are responsible for recruiting the Polycombs. To begin to test this, they inserted a 4kb sequence, encompassing the entire human HBA gene, into the orthologous position in the mouse locus and show by chromatin IP that it is capable of recruiting Polycomb group proteins. However, unfortunately they do not insert either the DNA fragments up or down stream of the CpG island within this 4kb fragment, to support their claim that the CpG island is required. They go on to mine previously published genome wide datasets of H3K27me3 enrichment profiles in human and mouse ES cells. In doing this, they nicely establish that the Polycomb target genes in human ES cells that are not enriched with H3K27me3 in mouse ES cells, contain CpG islands in human, but not in mouse cells. Finally, they also show, albeit on just two genes, that the loss of Dnmt1 or Dnmt3b/3a is correlated with a 2-fold increase in H3K27me3, but no change in H3K4me3.

Overall, the methods and approach of this paper is original, employing a powerful method of inserting DNA fragments into mouse ES cells in a controlled number of copies and performing a comparative study of chromatin states. However, the work relies heavily on data mining of previously published work and does not go far enough to warrant publication in EMBO J at this point.

Major points:

1. In Figure 3, the authors should test if the regions of DNA up stream and downstream of the CpG island are capable of recruiting Polycomb group proteins. If their hypothesis is correct, only the middle CpG island region should be capable of recruiting Polycombs.

This is a good suggestion, however we feel that it is beyond the scope of the current work since our specific hypothesis is that the density of unmethylated CpG dinucleotides is the key factor and this is tested by the insertion of a negative control fragment (the HBB gene) that has low CpG content and does not recruit Polycomb group proteins. In addition the evidence for unmethylated CpG dinucleotides is greatly strengthened by the new chip seq data presented for DNMT3ab -/- deficient
ES cells (Figure 7).

2. Panels I-L of Figure 6 demonstrate that two genes, Ebf3 and Zfp423, have a two fold greater enrichment of H3K27me3 on their gene loci in Dnmt1 null and Dnmt3a/3b null cells. While this is not a very significant change, it could support their hypothesis that loss of methylation on CpG islands leads to an increased presence of Polycomb group proteins. It would significantly strengthen their argument if they were to perform a genome wide CHIP-SEQ analysis of H3K27me3 in these WT versus knock cells.

We agree that the original qPCR data was not very convincing as to the generality of this phenomenon. To address this we have focused our analysis on the Dnmt3a/b -/- cells and performed genome wide chip sequencing for H3K27me3 in these cells and in wild type controls. As described in the text, we identified numerous additional examples of CpG islands which exhibit robust de novo polycomb recruitment in these knockout cells (Figure 7). For a number of examples we demonstrated by bisulphite sequencing that these sites are methylated in wild type ES cells and become demethylated in the knockout cells (Figure S6). Data from the Dnmt1 -/- cell line does not add to the conclusions and has been removed from the manuscript.

Minor points:

1) Several figures in this paper present CHIP analysis that incorrectly use "relative enrichment" as opposed to the accepted standard of "percentage input". These figures are Figure 1(C-F), Figure 2 (B, C, D), Figure 3 (B), Figure 4 (D, C, E), Figure 6 (I, J, K, L). These figures should be re-made to present the positive and negative controls, e.g. GATA6 and Actin.

As requested all figures have been remade to display chip data as a percentage of input. Positive and negative control points have been added to Figure 1 and Figure 4.

2) Legend to Figure 1 (B): It is not clear from the legend how the region was replaced. Wallace et al, (2007) should be referenced, as in the main text, or the method explicitly stated.

I have inserted the Wallace reference into the legend to Figure 1B.

3) Results section, page 7, paragraph two: no reference or evidence included for FERD3L being associated with bivalent chromatin.

Publically available ChIP seq data for the FERD3L gene are illustrated (Figure S1E).

4) Figure 3A: a scale indicating the fragment size would add to this panel.

A scale indicating fragment size has been added.

5) Figure 5A: It would help the reader if the key genes in this paper were highlighted to the right on panel A, e.g. MYO1G, CLEC4G, MYF6 and GATA6.

I have highlighted the position of these genes.

6) Figure Legend 6(I-L): Two minor omissions from this figure legend. Firstly, the positive controls, Gata6 (for the H3K27me3 ChIP) and Beta Actin (for the H3K4me3 ChIP) are not mentioned. Secondly, the cell type should be stated in the legend.

This qPCR data has been replaced by chip-seq data and moved to Figure 7.

7) The y-axis lines are missing in the ChIP-SEQ data in Figure Panels 3C, 5D and 6A.

Y axes lines have been added for all displayed chip-seq data (Figures 1, 3, 5,6, S1, S5).

8) Figure 6A, although it is mentioned in the main text that the intragenic RHBDF1 gene region is methylated in both human ES and somatic cells, this data should be presented in the Figure itself and also described in the Figure legend. It would also be appropriate detail the cell type that the
Cytosine methylation data was obtained from.

The cytosine methylation track in Figure 6A is derived from genome-wide bisulfite-sequencing in human ES cells (Lister et al, 2009). Data for other somatic cell types was presented in a previous publication from our laboratory (Vyas et al, 1992) in which this CGI is referred to as ‘89-91’. The methylation status was assessed by southern blot. This CGI was shown to be unmethylated in adult sperm and in fetal brain, partially methylated in fetal muscle and methylated in erythroid cell lines, peripheral blood mononuclear cells, B lymphocytes and placenta.

Response to Referee 3

"An interspecies analysis reveals determinants of the bivalent chromatin state" Lynch et al. investigates the bivalent chromatin state using a humanized mouse model. They show that differences in bivalency at the HBA2 and FERD3L loci between human and mouse ES cells are dominated by differences in regulatory sequences. Analysis of two non-overlapping regions of the HBA2 locus reveals that the bivalent state appears to be, at least in part, independently encoded in both fragments, suggesting direct recruitment of PcG proteins to multiple sites of a bivalent domain rather than spreading of said domain from an initial region of recruitment. Furthermore, the authors show that PRC2 occupancy is negatively correlated with transcriptional activity of HBA2. They computationally re-analyze public data to see whether CpG island erosion is associated with loss of H3K27me3 occupancy. The authors find that CpG islands that are marked with H3K4me3 or H3K27me3 are strongly biased towards unmethylated CpG islands and that demethylation of two analyzed CpG islands is accompanied with an increase in H3K27me3 but not H3K4me3 occupancy.

Given the important role of PcG proteins in developmental gene regulation and the limited knowledge of its regulation, the investigated topic is potentially of high interest. The recombination mediated cassette exchange approach, which controls location and copy number of inserted human DNA fragments into a mouse background, is elegant and well suited for the detailed analysis of gene regulation and chromatin state.

My single biggest criticism is that this elegant tool is not used to its fullest potential, because the authors consistently take single-locus approaches over much more persuasive global analyses. For instance, Figs 1-4 are in essence a series of bar-charts for about 4 site-specific ChIP reactions. It is difficult to extrapolate to the entire 120+ kilobase locus they are interrogating from these few data points. For an EMBO Journal calibre paper, I would hope to see the opposite: namely, ChIP-seq data in Figures 1-4, confirmed by site-specific analysis. Genome-wide data can also much more scientifically compelling and persuasive, because the transcriptional state of the entire human domain inserted into mouse would be presented within the figure, and the texture of the data would be clearer. Bar charts are rarely, if ever, intuitive.

Overall, most of the global conclusions are inferred from a limited number of experimental points, and thus their actual generality is difficult to evaluate. The paper would be tremendously strengthened if these elegant mouse constructs were interrogated using genome-wide tools, and the literature datasets were tested using genome-wide experiments.

We agree that genome-wide data can be more compelling than a limited number of qPCR points and it is for this reason that we have presented the genome-wide data for Ezh2 in Figure 3C. However, in figures 2-4, we are inserting short human DNA fragments (2-4kb in size) into the mouse locus and these are covered by multiple qPCR points. We do not believe that the generation of genome-wide data sets would add significantly to the interpretation of these experiments. To address the referee’s concerns that the number of qPCR points is too limited we have expanded this analysis with additional qPCR probes for the mouse alpha globin promoter (Figure 1C-F), three additional qPCR points throughout the endogenous mouse alpha globin locus (Figure 2,3) and additional qPCR points within fragment ii (Figure 2,3) and within the HBB fragment (Figure 2).

Specific comments:

1. The authors state: "(H)owever the human alpha globin locus is associated with prominent sites of
PcG recruitment and chromatin bivalency in pluripotent cells whereas the corresponding mouse locus is not (Garrick et al, 2008). Unfortunately, the previous publication by the authors (Garrick et al, 2008) does not provide this evidence. The only relevant comment I found in the 2008 paper is the claim in the discussion stating: “It is of interest that, unlike the human a-globin genes, we do not detect binding of the PRC2 complex and H3K27me3 modification at the orthologous mouse a-globin genes in nonexpressing cell types (data not shown).” When I realized that their prior publication was not about interspecies comparisons at all (in contrast to how they pitched Garrick 2008 in the intro), I felt like they needed to show in detail the human versus mouse differences before (or perhaps, within?) exploring it within this current Lynch et al manuscript.

To address this concern, we have added a more detailed comparison of chromatin state at these loci (Figure S1A-D).

2. The authors use ChIP-seq data from Ku et al. to support the difference in bivalency between the human and mouse HBA locus. The Ku data are presented in such a large window that it is difficult to evaluate in detail the chromatin state at the a-globin locus. It would be more convincing to also include data for other appropriate chromatin marks that are available from the same publication (H3K4me3, H3K27me3, etc.).

These data have been added to display the chromatin state across the entire locus and zoomed in at the HBA2 gene in Figure S1A-D.

Indeed, when I inspected the HBA locus in more detail using UCSC, although the locus shows chromatin marks consistent with bivalency, the chromatin state of HBA1 and HBA2 can not be cleanly evaluated due to alignability issues. I would like to see a more explicit and detailed description of how the a-globin locices align in mouse and human, maybe in supplemental. This issue needs to be carefully addressed when ChIP-seq data are used to evaluate occupancy patterns at repetitive genomic locations.

In both the mouse and human genomes the alpha globin genes are arranged in duplicated homology blocs (illustrated in Fig S1A,B). The default behaviour for most read mapping algorithms is to discard reads that match more than once in the genome. It is for this reason that the referee could not see signal at the alpha globin genes on the UCSC tracks. All of the chip seq data displayed in this manuscript have been remapped using Bowtie with setting m=2, which permits up to two exact matches of a read within the genome (note that this was also the case in the previous version of the manuscript).

2. The ChIP-qPCR data in Figures 1 C-F and 4 B-E lack negative control regions, as is shown for example in Figure 2 B-D.

Negative and positive control points have been added to these figures.

3. Throughout the manuscript the authors do not attempt to normalize the levels of H3K4me3 and H3K27me3 to bulk H3 occupancy. The observed changes in H3K4me3 and H3K27me3 occupancy could be explained by an increase in methylation levels, or alternatively by an increase in general nucleosome occupancy at a given location.

We do not believe that changes in histone occupancy can be responsible for the observed differences in H3K4me3 and H3K27me3 in Figures 2,3 and 4 since the changes are reciprocal ñ where H3K27me3 is reduced H3K4me3 is increased and vice versa. Nevertheless, the absence of H3K4me3 and H3K27me3 at the HBB fragment in Figure 2 could, in principle, reflect a decrease in histone occupancy. To address this we performed chip for unmodified H3 for cell lines containing the HBA2, HBB and FERD3L fragments and found no significant differences in histone occupancy (Figure S1G-I). In addition it is noted that changes in H3K27me3 are correlated with changes in PRC2 binding.

4. From the experiments presented in Figure 3 the authors conclude that the bivalent chromatin state is encoded multiple times within the HBA2 locus. This is an important point of the paper, and could be a major finding, but the experimental design has certain limitations:
i) All fragments analysed overlap a specific CpG island. Thus the presented data would also be consistent with recruitment of PRC2 to the CpG island and spreading from there. In other words; would fragment (ii) show bivalency when significantly shorter and thus not overlapping the CpG island at the HBA2 promoter?

This would be an interesting experiment, however we feel that it is beyond the scope of the current work since our specific hypothesis is that the density of unmethylated CpG dinucleotides is the key factor and this is tested by the insertion of a negative control fragment (the HBB gene) that has low CpG content and does not recruit Polycomb group proteins. In addition the evidence for unmethylated CpG dinucleotides is greatly strengthened by the new chip seq data presented for DNMT3ab-/- deficient ES cells (Figure 7).

ii) Fragment (ii) was analyzed at only a single point. This point showed marginal-to-no enrichment of Ezh2 and H3K27me3 in Figure 2B but the conclusions from this paragraph rest on one data point. This is related to the advantage of using ChIP-seq as primary data, versus confirmation data, above.

We have designed an additional qPCR probe within fragment (ii) (numbered 5 in Figure 2) and data for this point shown in Figure 2B and Figure 3B confirms enrichment above background for Ezh2 and H3K27me3.

The ChIP-seq data in Figure 3 C are presented without an input control and lack experiments for H3K27me3 and H3K4me3. Also the alignability issues as mentioned in point 1 have to be addressed.

An input track has been added to the figure. The alignability issues are discussed above.

5. From the experiment in Figure 4 the authors conclude: "There is a linear, inverse relationship between H3K27me3 and H3K4me3 (Figure 4C-F) with the highest level of PRC2 recruitment and lowest level of H3K4me3 seen when the activating sequences associated with the promoter were deleted. Complete clearing of PRC2 and H3K27me3 occurred in the presence of the MC1 promoter and an intermediate level of H3K4me3 and H3K27me3 was observed for the wild type construct." To assume a linear relationship from only three data points is perhaps overly confident.

We agree that it is premature to conclude a linear relationship from three data points and this has been changed to "inverse relationship" in the text.

By showing relative enrichment, rather than absolute enrichment relative to a genomic control known to be not enriched, it is not possible to evaluate if there is "complete clearing of PRC2 and H3K27me3". In general, all of the presented ChIP-qPCR results would be easier to evaluate and compare if they were presented as absolute rather than relative enrichment.

All of the figures displaying ChIP-qPCR data are now presented as percentage of input material. Positive and negative control points have been added to Figure 4C to permit the conclusion that complete clearing has occurred.

6. Their analysis of CpG data from other groups (mostly Figure 5) was not strong. Singling out <10% of the CpG islands that vary between human and mouse is not convincing, nor were the three, hand-selected examples where the CpG changes correspond to changes in the methylation status (Fig 5D-F).

This is by its nature a correlative analysis, nonetheless a clear trend is apparent where loss of CpG density in the mouse compared to the human genome is associated with loss of H3K27me3 recruitment.

7. The number of loci (2) analysed in Figure 6G to L is too small. These experiments need to be repeated with a larger sample.

We agree that the original qPCR data was not very convincing as to the generality of this phenomenon. To address this we have focused our analysis on the Dnmt3a/b-/- cells and performed
genome wide chip sequencing for H3K27me3 in these cells and in wild type controls. As described in the text, we identified numerous additional examples of CpG islands which exhibit robust de novo polycomb recruitment in these knockout cells (Figure 7). For a number of examples we demonstrated by bisulfite sequencing that these sites are methylated in wild type ES cells and become demethylated in the knockout cells (Figure S6). Data from the Dnmt1 −/− cell line does not add to the conclusions and has been removed from the manuscript.

8. How do the authors explain that there is no increase in H3K4me3 occupancy in Figures 6J and L since H3K27me3 alone occupies only a minority of CpG islands in ES cells (Ku et al Figure 4A).

In general, the sequences responsible for recruitment of H3K4 methyltransferases are poorly understood. Our hypothesis is that a high density of unmethylated CpG dinucleotides is sufficient for Polycomb recruitment but that additional sequence elements are required for the efficient recruitment of H3K4 methyltransferases. We note that a large proportion of the de novo sites of PRC2 recruitment that we observe in Figure 7 correspond to gene exons. In this genomic context, the CpG density is likely to reflect codon preference rather than promoter function and therefore additional sequence elements required for H3K4me3 methyltransferase recruitment may be absent even when these sequences lose DNA methylation in DNMT 3a/b −/− cells.

9. The numbers indicating the scale of the displayed ChIP-seq data in Figure 6A are very small. Also data sets from the same factors in different cell types should be shown at the same scale and normalized for sequencing depth in order to be comparable.

All chip seq data is now normalized for the number of mapped reads with the scale indicating mapped reads per 10 million. Data sets in Figure 6A are now shown on the same scales as requested.

Minor points:

1. On page 9 "Fragment 1" should be Fragment (i).
   Corrected

2. 'Salk Institute' or 'Broad Institute' are not appropriate references. The authors should provide references to the relevant papers.
   Corrected

3. Scales missing for heatmaps in Figure 5A.
   I have added these scales

4. All primer and probe sequences should be in a supplemental file.
   I have added these into the supplementary data

5. In Figures 1 and 3 "Ku et al data" should be "Ku et al. data".
   Corrected

2nd Editorial Decision 20 September 2011

Thank you for submitting your revised manuscript for consideration by The EMBO Journal, and sorry for the delay in its re-evaluation. All three of the original referees have now reviewed it once more, and I am happy to inform you that they consider the majority of specific criticisms satisfactorily addressed. Referee 3 maintains his/her conceptual criticism about general ChIP-seq data requirement, but as discussed previously (and as also confirmed by referee 1) I am not going to insist on such a requirement given the specific nature of the current study and the experimental approach of genomic cassette swapping. At the same time, there are a number of minor/specific points remaining to be addressed in the new reports of referees 1 and 2, which I would kindly like to ask you to address through a final round of minor revision. The only more significant open issue therefore remains referee 2's point 1, as this referee is not fully satisfied with your current response. While I am not sure whether additional experiments would be absolutely required to address this concern, I would appreciate if you could carefully consider this point and get back to me with thoughts on how to best clarify and reconcile this issue.
I am thus returning the study to you once more with the request to consider and respond/address referee 1 and 2's remaining points. When sending a re-revised manuscript, please make sure to include manuscript and supplementary files without mark-up's for changes. I look forward to hearing from you.

Sincerely,

Editor
The EMBO Journal

---

REFeree REPORTS:

Referee #1 (Remarks to the Author):

Comments for Lynch et al

I am very satisfied with the authors amendments.

Figure 7
I really liked the authors additional data studying Dnmt3a/b KO cells. However, it would be nice if they could show changes in H3K27me3 globally by plotting H3K27me3 read density for WT vs Dnmt3a/b cells for all, methylated (>80%) and unmethylated (<5%) CpG islands (similar to Fig 6B-E). This would give the reader a better perspective as to what proportion of CpG islands become de novo H3K27me3 modified.

I noticed that in Figure 6 legend the authors write "DNA methylation prevents polycomb recruitment in pluripotent cells" and Figure 7 legend they write "De novo PRC2 recruitment in Dnmt3a/b -/- ES cells" and in the final paragraph of the discussion they write "we observed numerous examples of de novo PcG recruitment". In the last part of the abstract they also say "multiple sites of de novo Polycomb recruitment in .". From what I can see in all these cases the authors are studying H3K27me3 marks and not PcG proteins. I appreciate that the H3K27me3 mark was probably applied by PRC2 and will probably recruit other PcG proteins but this is not what has been tested, so I wonder if the text could be amended.

Reviewer 3 suggested the authors should make greater use of ChIP-seq for figure 1-4. Although I agree sequencing (or ChIP-chip) data can provide a more complete picture, in this case because the human gene fragments only constitute a very small portion of the genome (approx 4 kb) this experiment would essentially just be analysing these marks in WT mouse cells. I also feel a detailed ChIP-seq analysis around the human locus would add little additional information to this story and the authors do already show significant EZH2 recruitment to the human Hba2 fragment compared to the adjacent mouse Hba1 locus (Fig 3c).

Minor point
Figure 6A - Type H3K2me3 should be H3K27me3

Referee #2 (Remarks to the Author):

Comments on the authors' responses:

> Major points:

> 1. In Figure 3, the authors should test if the regions of DNA up stream and downstream of the CpG island are capable of recruiting Polycomb group proteins. If their hypothesis is correct, only the middle CpG island region should be capable of recruiting Polycombs.
This is a good suggestion, however we feel that it is beyond the scope of the current work since our specific hypothesis is that the density of unmethylated CpG dinucleotides is the key factor and this is tested by the insertion of a negative control fragment (the HBB gene) that has low CpG content and does not recruit Polycomb group proteins. In addition the evidence for unmethylated CpG dinucleotides is greatly strengthened by the new chip seq data presented for DNMT3ab −/− deficient ES cells (Figure 7).

This answer does not satisfy the core criticism of this figure (Figure 3), which is that all three shortened fragments (II, III and IV) still contain parts of the CpG island to varying amounts. It is therefore not at all surprising that all three fragments STILL recruit Polycombs. This reviewer's argument is that it would be much more satisfying to see an experiment where the authors tested the up- and down-stream elements which do not overlap the CpG island. One would expect to see that these fragments would not recruit Polycombs if the author's hypothesis is correct. Furthermore, their argument that the "control fragment" (the HBB gene) has no Polycomb recruitment could be due to reasons completely unrelated to CpG content.

2. Panels I-L of Figure 6 demonstrate that two genes, Ebf3 and Zfp423, have a two fold greater enrichment of H3K27me3 on their gene loci in Dnmt1 null and Dnmt3a/3b null cells. While this is not a very significant change, it could support their hypothesis that loss of methylation on CpG islands leads to an increased presence of Polycomb group proteins. It would significantly strengthen their argument if they were to perform a genome wide CHIP-SEQ analysis of H3K27me3 in these WT versus knock cells.

We agree that the original qPCR data was not very convincing as to the generality of this phenomenon. To address this we have focused our analysis on the Dnmt3a/b −/− cells and performed genome wide chip sequencing for H3K27me3 in these cells and in wild type controls. As described in the text, we identified numerous additional examples of CpG islands which exhibit robust de novo polycomb recruitment in these knockout cells (Figure 7). For a number of examples we demonstrated by bisulphite sequencing that these sites are methylated in wild type ES cells and become demethylated in the knockout cells (Figure S6). Data from the Dnmt1 −/− cell line does not add to the conclusions and has been removed from the manuscript.

It certainly helps that the authors have now performed ChIP sequencing of H3K27me3 in Dnmt3a/b null cells. However, why do the authors select a limited cohort of genes to present without first presenting an overview of all genes? It would be good to have an objective overview of ALL CpG islands and a correlation between their respective gain/loss of Polycomb group proteins.

A new minor point on the revised Figure 7 is that the scale of the Y axes on all of the ChIP-SEQ plots should be the same between WT (5) and KO (3).

Minor points:

1) Several figures in this paper present CHIP analysis that incorrectly use "relative enrichment" as opposed to the accepted standard of "percentage input". These figures are Figure 1(C-F), Figure 2 (B, C, D), Figure 3 (B), Figure 4 (D, C, E), Figure 6 (I, J, K, L).

These figures should be re-made to present the positive and negative controls, e.g. GATA6 and Actin.

As requested all figures have been remade to display chip data as a percentage of input. Positive and negative control points have been added to Figure 1 and Figure 4.

This is now fine.

2) Legend to Figure 1 (B): It is not clear from the legend how the region was replaced. Wallace et al, (2007) should be referenced, as in the main text, or the method explicitly stated.

I have inserted the Wallace reference into the legend to Figure 1B.

This is now fine.
> 3) Results section, page 7, paragraph two: no reference or evidence included for FERD3L being associated with bivalent chromatin.

>> Publically available ChIP seq data for the FERD3L gene are illustrated (Figure S1E).

. This is now fine.

> 4) Figure 3A: a scale indicating the fragment size would add to this panel.

>> A scale indicating fragment size has been added.

. This is now fine.

> 5) Figure 5A: It would help the reader if the key genes in this paper were highlighted to the right on panel A, e.g. MYO1G, CLEC4G, MYF6 and GATA6.

>> I have highlighted the position of these genes.

. This is now fine.

> 6) Figure Legend 6(I-L): Two minor omissions from this figure legend. Firstly, the positive controls, Gata6 (for the H3K27me3 ChIP) and Beta Actin (for the H3K4me3 ChIP) are not mentioned. Secondly, the cell type should be stated in the legend.

>> This qPCR data has been replaced by chip-seq data and moved to Figure 7.

. This is now fine.

> 7) The y-axis lines are missing in the ChIP-SEQ data in Figure Panels 3C, 5D and 6A.

>> Y axes lines have been added for all displayed chip-seq data (Figures 1, 3, 5, 6, S1, S5).

. This is now fine.

> 8) Figure 6A, although it is mentioned in the main text that the intragenic RHBDF1 gene region is methylated in both human ES and somatic cells, this data should be presented in the Figure itself and also described in the Figure legend. It would also be appropriate detail the cell type that the Cytosine methylation data was obtained from.

>> The cytosine methylation track in Figure 6A is derived from genome-wide bisulfite-sequencing in human ES cells (Lister et al, 2009). Data for other somatic cell types was presented in a previous publication from our laboratory (Vyas et al, 1992) in which this CGI is referred to as '89/-91'. The methylation status was assessed by southern blot. This CGI was shown to be unmethylated in adult sperm and in fetal brain, partially methylated in fetal muscle and methylated in erythroid cell lines, peripheral blood mononuclear cells, B lymphocytes and placenta.

. The suggestions have not yet been adequately addressed in the revised figure. We propose labeling the Cytosine methylation track shown with the cell type, in this case "ES cells". Furthermore, it would be preferable if they can also add a similar track for the Erythroblast cell line.

Referee #3 (Remarks to the Author):

The authors addressed many of the minor comments I provided, but rejected my primary concern regarding the suitability of site-specific analyses for the kinds of conclusions they are making. I do appreciate that there are a number of technical improvements to this newer version, but this does not cross the threshold for publication.

Nowadays HTP sequencing is a routine end-point for ChIP assays across the large majority of
transcriptional research labs; site-specific analysis is no longer sufficient for high-quality journals. To publish in EMBO Journal a study of this nature demands global data (e.g. ChIP-seq of all their experiments), which has been carefully mapped and analyzed, and only validated by experiments such as they principally report here.

Therefore, I have to recommend (with regret, given how much I like their model system and approach) that this paper be forwarded to a more specialized journal.

2nd Revision - authors’ response 05 October 2011

Many thanks for the opportunity to resubmit our revised manuscript to be considered for publication in the EMBO Journal. We were very pleased to see that our revisions have addressed the referees’ original criticisms. We think that their suggested changes have significantly improved the manuscript.

Originally, two referees asked for more examples (than the two we provided) showing that when methylated CpG islands become unmethylated they recruit Polycomb (PcG). We addressed this by identifying CpG islands that are bound by PcG in Dnmt3a/b knockout ES cells (presumably unmethylated) but not in the corresponding wild type ES cells (presumably methylated). We then carefully determined the status of cytosine methylation at each of these loci using bisulfite PCR followed by cloning of the PCR product and sequencing of multiple clones for wild type and mutant ES cells. This allowed us to provide the referees with 8 new examples of this phenomenon, consolidating our original conclusion. The new request for genome-wide analysis of this effect would require us to sequence bisulfite-converted genomic DNA from these cell lines to 20-30 fold coverage, which remains prohibitively expensive and is (in our opinion) beyond the scope of this study.

The second outstanding issue appears to be reviewer #2’s concern that sequences upstream and downstream of the annotated CpG islands, which do not normally recruit PcG in their natural chromosomal position (Figure S1C), might newly acquire this property when inserted into the RMCE system. Clearly, the control from the beta globin gene that we used (which also does not bind PcG in vivo) does not acquire this property. So the referee is asking us to test whether some fragments (including the alpha CpG island and its flanking sequences), but not others acquire the ability to recruit PcG when placed in the RMCE system. This seems very unlikely given the supporting evidence we provide for recruitment via the CpG island itself. This includes:

(i) The greatest magnitude of PcG recruitment is localised in the central CpG-rich area of the HBA2 fragment and not the peripheral CpG-poor regions (Figure S1C)

(ii) Demethylation of constitutively methylated CpG islands is associated with the appearance of de novo genomic sites of the PcG-associated H3K27me3 modification (Figure 7)

(iii) CpG island erosion during evolution is associated with loss of PcG recruitment (Figure 5).

In addition to these lines of evidence, recently published genome-wide analyses have found most sites of PcG recruitment to be associated with CpG-rich sequences (Ku et al, 2008) and that CpG-rich sequences can be sufficient for PcG recruitment to ectopic genomic sites (Mendenhall et al, 2010).

The main new conclusion from Figure 3 therefore is not that CpG islands recruit PcG but that this property is redundantly encoded within such a region. The referee does not dispute that this conclusion is fully supported by the data presented and indeed finds it “unsurprising”. Clearly we do have the ability to perform this experiment but it would delay publication for about 6 months while we construct and analyse another RMCE clone. Given all of the evidence supporting the conclusion that CpG islands provide a platform for the recruitment of PcG, in our view the time required to
generate and analyse this control would not be justified by the potential scientific benefit. Our concern is also that a long delay in publication might significantly compromise the impact of our paper in a rapidly moving field.

Other minor changes to the manuscript are detailed in our point by point response to the reviewers. We hope that with these changes you now consider our manuscript suitable for publication in the EMBO Journal. If you have any further queries please do not hesitate to contact me.

Response to Referee 1

I am very satisfied with the authors amendments.

Figure 7
I really liked the authors additional data studying Dnmt3a/b KO cells. However, it would be nice if they could show changes in H3K27me3 globally by plotting H3K27me3 read density for WT vs Dnmt3a/b cells for all, methylated (>80%) and unmethylated (<5%) CpG islands (similar to Fig 6B-E). This would give the reader a better perspective as to what proportion of CpG islands become de novo H3K27me3 modified.

Originally, two referees asked for more examples (than the two we provided) showing that when methylated CpG islands become unmethylated they recruit Polycomb (PcG). We addressed this by identifying CpG islands that are bound by PcG in Dnmt3a/b knockout ES cells (presumably unmethylated) but not in the corresponding wild type ES cells (presumably methylated). We then carefully determined the status of cytosine methylation at each of these loci using bisulfite PCR followed by cloning of the PCR product and sequencing of multiple clones for wild type and mutant ES cells. This allowed us to provide the referees with 8 new examples of this phenomenon, consolidating our original conclusion. We would have liked to extend this to a genome-wide analysis, as shown for human ES cells in Figure 6, however in contrast to the situation for human ES cells, high resolution bisulfite sequencing maps of methylation are available for neither wild type nor Dnmt3a/b deficient mouse ES cells. The generation of this data would require us to sequence bisulfite-converted genomic DNA from these cell lines to 20-30 fold coverage, which remains prohibitively expensive and is (in our opinion) beyond the scope of this study. In the absence of this data, global changes in H3K27me3 are very difficult to interpret as the methylation status of any given CpG island is unknown in either the mutant or the wild type ES cells.

I noticed that in Figure 6 legend the authors write "DNA methylation prevents polycomb recruitment in pluripotent cells" and Figure 7 legend they write "De novo PRC2 recruitment in Dnmt3a/b-/- ES cells" and in the final paragraph of the discussion they write "we observed numerous examples of de novo PcG recruitment". In the last part of the abstract they also say "multiple sites of de novo Polycomb recruitment in". From what I can see in all these cases the authors are studying H3K27me3 marks and not PcG proteins. I appreciate that the H3K27me3 mark was probably applied by PRC2 and will probably recruit other PcG proteins but this is not what has been tested, so I wonder if the text could be amended.

The Referee correctly points out that we have employed the H3K27me3 histone modification as a surrogate for PcG recruitment. To clarify this, the legends for Figures 6 and 7 and the abstract and discussion have been modified as suggested.

Reviewer 3 suggested the authors should make greater use of ChIP-seq for figure 1-4. Although I agree sequencing (or ChIP-chip) data can provide a more complete picture, in this case because the human gene fragments only constitute a very small portion of the genome (approx 4 kb) this experiment would essentially just be analysing these marks in WT mouse cells. I also feel a detailed ChIP-seq analysis around the human locus would add little additional information to this story and the authors do already show significant EZH2 recruitment to the human Hba2 fragment compared to the adjacent mouse Hba1 locus (Fig 3c).

Minor point
Figure 6A - Type H3K2me3 should be H3K27me3
Corrected
Response to Referee 2

Major points:

1. In Figure 3, the authors should test if the regions of DNA upstream and downstream of the CpG island are capable of recruiting Polycomb group proteins. If their hypothesis is correct, only the middle CpG island region should be capable of recruiting Polycombs.

This is a good suggestion, however we feel that it is beyond the scope of the current work since our specific hypothesis is that the density of unmethylated CpG dinucleotides is the key factor and this is tested by the insertion of a negative control fragment (the HBB gene) that has low CpG content and does not recruit Polycomb group proteins. In addition the evidence for unmethylated CpG dinucleotides is greatly strengthened by the new chip seq data presented for DNMT3ab-/- deficient ES cells (Figure 7).

This answer does not satisfy the core criticism of this figure (Figure 3), which is that all three shortened fragments (II, III and IV) still contain parts of the CpG island to varying amounts. It is therefore not at all surprising that all three fragments STILL recruit Polycombs. This reviewers argument is that it would be much more satisfying to see an experiment where the authors tested the up- and down-stream elements which do not overlap the CpG island. One would expect to see that these fragments would not recruit Polycombs if the author's hypothesis is correct. Furthermore, their argument that the "control fragment" (the HBB gene) has no Polycomb recruitment could be due to reasons completely unrelated to CpG content.

Referee 2 is concerned that sequences upstream and downstream of the annotated CpG islands, which do not normally recruit PcG in their natural chromosomal position (Figure S1C), might newly acquire this property when inserted into the RMCE system. Clearly, the control from the beta globin gene that we used (which also does not bind PcG in vivo) does not acquire this property. So the referee is asking us to test whether some fragments (including the alpha CpG island and its flanking sequences), but not others acquire the ability to recruit PcG when placed in the RMCE system. This seems very unlikely given the supporting evidence we provide for recruitment via the CpG island itself. This includes:-

(i) The greatest magnitude of PcG recruitment is localised in the central CpG-rich area of the HBA2 fragment and not the peripheral CpG-poor regions (Figure S1C)

(ii) Demethylation of constitutively methylated CpG islands is associated with the appearance of de novo genomic sites of the PcG-associated H3K27me3 modification (Figure 7)

(iii) CpG island erosion during evolution is associated with loss of PcG recruitment (Figure 5).

In addition to these lines of evidence, recently published genome-wide analyses have found most sites of PcG recruitment to be associated with CpG-rich sequences (Ku et al, 2008) and that CpG-rich sequences can be sufficient for PcG recruitment to ectopic genomic sites (Mendenhall et al, 2010).

The main new conclusion from Figure 3 therefore is not that CpG islands recruit PcG but that this property is redundantly encoded within such a region. The referee does not dispute that this conclusion is fully supported by the data presented and indeed finds it unsurprisingi. Clearly we do have the ability to perform this experiment but it would delay publication for about 6 months while we construct and analyse another RMCE clone. Given all of the evidence supporting the conclusion that CpG islands provide a platform for the recruitment of PcG, in our view the time required to generate and analyse this control would not be justified by the potential scientific benefit.

2. Panels I-L of Figure 6 demonstrate that two genes, Ebf3 and Zfp423, have a two fold greater enrichment of H3K27me3 on their gene loci in Dnmt1 null and Dnmt3a/3b null cells. While this is not a very significant change, it could support their hypothesis that loss of methylation on CpG islands leads to an increased presence of Polycomb group proteins. It would significantly strengthen
their argument if they were to perform a genome wide CHIP-SEQ analysis of H3K27me3 in these WT versus knock cells.

We agree that the original qPCR data was not very convincing as to the generality of this phenomenon. To address this we have focused our analysis on the Dnmt3a/b-/- cells and performed genome wide chip sequencing for H3K27me3 in these cells and in wild type controls. As described in the text, we identified numerous additional examples of CpG islands which exhibit robust de novo polycomb recruitment in these knockout cells (Figure 7). For a number of examples we demonstrated by bisulphite sequencing that these sites are methylated in wild type ES cells and become demethylated in the knockout cells (Figure S6). Data from the Dnmt1-/- cell line does not add to the conclusions and has been removed from the manuscript.

- It certainly helps that the authors have now performed ChIP sequencing of H3K27me3 in Dnmt3a/b null cells. However, why do the authors select a limited cohort of genes to present without first presenting an overview of all genes? It would be good to have an objective overview of ALL CpG islands and a correlation between their respective gain/loss of Polycomb group proteins.

Originally, two referees asked for more examples (than the two we provided) showing that when methylated CpG islands become unmethylated they recruit Polycomb (PcG). We addressed this by identifying CpG islands that are bound by PcG in Dnmt3a/b knockout ES cells (presumably unmethylated) but not in the corresponding wild type ES cells (presumably methylated). We then carefully determined the status of cytosine methylation at each of these loci using bisulfite PCR followed by cloning of the PCR product and sequencing of multiple clones for wild type and mutant ES cells. This allowed us to provide the referees with 8 new examples of this phenomenon, consolidating our original conclusion. We would have liked to extend this to a genome-wide analysis, as shown for human ES cells in Figure 6, however in contrast to the situation for human ES cells, high resolution bisulfite sequencing maps of methylation are available for neither wild type nor Dnmt3a/b deficient mouse ES cells. The generation of this data would require us to sequence bisulfite-converted genomic DNA from these cell lines to 20-30 fold coverage, which remains prohibitively expensive and is (in our opinion) beyond the scope of this study. In the absence of this data, global changes in H3K27me3 are very difficult to interpret as the methylation status of any given CpG island is unknown in either the mutant or the wild type ES cells.

- A new minor point on the revised Figure 7 is that the scale of the Y axes on all of the Chip-SEQ plots should be the same between WT (5) and KO (3).

The ChIP-seq data for both WT and KO is normalized to show the read count per 10 million. Since the domain of binding appears to be broader for the KO (Figure 7, S5) and there are additional sites of binding in the KO cells, it is to be expected that reads will be distributed across a larger proportion of the genome and therefore the peak read density is lower at positive control points such as HoxA (Figure 7A) and Gata6 (Figure S5). For this reason we elected to display data for WT and KO cells on different scales. However please note that consistent scales are used throughout all panels in the Figure.

8) Figure 6A, although it is mentioned in the main text that the intragenic RHBDL1 gene region is methylated in both human ES and somatic cells, this data should be presented in the Figure itself and also described in the Figure legend. It would also be appropriate detail the cell type that the Cytosine methylation data was obtained from.

The cytosine methylation track in Figure 6A is derived from genome-wide bisulfite-sequencing in human ES cells (Lister et al, 2009). Data for other somatic cell types was presented in a previous publication from our laboratory (Vyas et al, 1992) in which this CGI is referred to as ‘-89/-91’. The methylation status was assessed by southern blot. This CGI was shown to be unmethylated in adult sperm and in fetal brain, partially methylated in fetal muscle and methylated in erythroid cell lines, peripheral blood mononuclear cells, B lymphocytes and placenta.

- The suggestions have not yet been adequately addressed in the revised figure. We propose labeling the Cytosine methylation track shown with the cell type, in this case "ES cells". Furthermore, it would be preferable if they can also add a similar track for the Erythroblast cell line.
As requested ‘ES cells’ has been added to the label for the cytosine methylation track. Generation of genome-wide bisulfite data requires the sequencing of bisulfite-converted genomic DNA to 20-30 fold coverage (as performed by Lister et al) and remains prohibitively expensive for most applications. Consequently we consider that the generation of such a data set for the erythroid cells is beyond the scope of the current work.

Response to Referee 3

*Nowadays HTP sequencing is a routine end-point for ChIP assays across the large majority of transcriptional research labs; site-specific analysis is no longer sufficient for high-quality journals. To publish in EMBO Journal a study of this nature demands global data (e.g. ChIP-seq of all their experiments), which has been carefully mapped and analyzed, and only validated by experiments such as they principally report here.*

We agree that in general ChIP-seq data is preferable to qPCR data and this is the routine practice in our laboratory, however we concur with referee 1 that for the series of experiments described in Figures 1-4 the human gene fragment constitutes a very small proportion of the genome and therefore the generation of global chip-seq data for the analysis of just this tiny part of the genome is not warranted or appropriate.