

A stepwise mode of TGF β -SMAD signaling and DNA methylation regulates naïve-to-primed pluripotency and differentiation

Corresponding Author: Dr Qiong Wang

Parts of this Peer Review File have been redacted as indicated to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

In their manuscript “Dnmt3b mediates TGF β -Smad2/3 action during epiblast priming from mouse embryonic stem cells independently of Smad4,” Zhao, Yu, Shi, Ma et al. make a series of interesting connections between several major genomic and transcriptional/regulatory dynamics that accompany early mouse embryogenesis. Beginning with generation of Smad2/3 and Smad4 KO mouse embryonic stem cell lines, the authors identify several waves of Smad binding and activity, including a clear role for Smad2/3 in enabling the naïve to primed transition from conventional mouse ESCs to mouse Epiblast-like cells (EpiLCs), which is followed by an auxiliary/secondary role of Smad4 in inducing mesendodermal regulators such as Eomes and Wnt3. Then, the authors use these cells to identify context-specific interaction partners via IP-Mass Spec, including an interaction between the DNA methyltransferase DNMT3B and SMAD2/3 within Smad4 KO cells, an interaction that is not as apparent in WT EBs. The authors then develop and test a model that postulates an intermediate role for this interaction either prior to or instead of canonical SMAD2/3-4-based gene regulation.

Overall, this is an interesting manuscript with several notable findings related to the stepwise progression of early postimplantation development, including developmental signal-transduction (Nodal-based Smad signaling), naïve-to-primed changes in chromatin state (hypo to hypermethylated genomes), and differentiation (Smad4-facilitated primitive streak induction). However, these very nice connections within the early portion of the manuscript are somewhat unnaturally merged with a biochemical model that is much harder to address and, to me, relies on weaker data. Overall, I feel that rectifying the strengths of these earlier figures with the biochemical information in Figures 4 and 5 will be quite challenging, but offer some comments on how to possibly address them in greater detail below.

To me, the biochemical interactions between SMAD2/3 and DNMT3B could very well be indirect and simply enriched within Smad4 KO EBs because this line produces a different combination of less differentiated cell types than either WT or Smad2/3 KO EBs. A streamlined model would be that Smad2/3 signaling is dominant in the naïve to primed transition (evident in Figures 2 and 3), induces DNMT3B, and this interaction is captured by nature of their shared nuclear localization and high expression. Then, Smad4 participates in downstream differentiation of primed state cells into mesodermal or endodermal cell types, which involves a more specific interaction with SMAD2/3 and promotes downstream cell types where DNMT3B is less actively engaged in global genome remethylation. In other words, the apparent interaction between SMAD2/3 and DNMT3B could very well be because these different KO lines enrich for different cell types along the pluripotent to differentiated continuum, with Smad4 KO EBs consisting of more pre-differentiated epiblast-like and early primitive streak cells. Similarly, Dnmt3B KD and OE experiments could phenocopy various results by altering the methylation status of primed state cells, which will rely on this modification for viability.

Ultimately, disentangling such fast and heterogeneous dynamics using bulk epigenomic and biochemical assays would be very tricky to impossible, and I do not feel that the very interesting aspects of this manuscript should be over-ridden by these technical limitations. As an alternative, I would ask that the authors consider making these other potential explanations clear in the text as well as rewording the title, abstract and model figure (Figure 6) to prioritize the strongest aspects of their data, namely a stepwise series of events that implicates a transcriptional role of Smad2/3 in supporting DNMT3B induction,

enrichment for nuclear localization and potential interactions between Smad2/3 and DNMT3B, followed by formation of Smad2/3-4 heterotrimers that appear to support primitive streak induction.

The figures themselves could also be further streamlined with substantial data being brought into the supplementary data – Figures 1, 2 and 3 ultimately address very similar points and could be reduced to two figures total. The key details to highlight are the distinct transcriptional dynamics between either Smad2/3 vs Smad4 KO, validation that Smad2/3 is sufficient for EpiLC-induction, and ChIP-seq data demonstrating that Smad2/3 binding precedes Smad4 binding and includes Dnmt3b locus.

Figure 4 and 5 seem especially problematic because they assume the interactions between SMAD2/3 and DNMT3B are direct, exist as an alternative interaction than SMAD4, and speak to the combined actions of these factors at chromatin. Alternatively, I would suggest that Smad4 KO EBs are less (or differently) differentiated than WT, have higher genomic DNMT3 activity, and that DNMT3B co-immunoprecipitates with SMAD2/3 under stimulating conditions because they are both nuclear. To substantiate their claim, I feel the authors need to establish that the SMAD2/3-DNMT3B interaction is specific, generalizable (not an artifact of cell state composition) and occurs at chromatin:

- To address specificity, I would ask that the authors consider IP'ing other factors that become induced or localized to the nucleus during the Naïve to Primed transition, such as pERK or Otx2, to highlight that DNMT3B is not just very abundant during this transition and interacting non-specifically with other nuclear factors.
- To address generalizability, I would ask that the authors consider validating these interactions in an ectopic cell system, even something as straightforward as HEK293 cells.
- To address their activity or interaction at chromatin, I would ask that the authors consider fractionating their cell models to demonstrate that the SMAD2/3-DNMT3B interaction is specific and at chromatin, rather than generic interactions within the cytoplasm or nucleoplasm. They could also consider altering the stringency of their IP buffers.

To me, Figure 5 does not highlight much beyond what is already known about DNMT3B recruitment or expression dynamics and doesn't show much convincing evidence that DNMT3B or SMAD2/3 are colocalized. If anything, Figure 5E fairly confidently establishes that they are enriched at different genomic features. Most comments about differential activity/behavior at the Fgf5 promoter is based on very weak signal of a single region. If the authors are motivated to showcase the degree to which these factors co-occupy similar regions of the genome, they should consider a genome-wide approach (DNMT3B ChIP-seq in their WT and KO cell types). Odds are fairly good though that these will largely confirm that these factors go to different genomic regions. Should the authors attempt a more genome-wide investigation, I would also ask that they include DNMT3A as it can be more clearly associated with other genomic features beyond gene bodies.

Furthermore, the inability to rescue delayed developmental defects with anything other than the full, catalytically active DNMT3B would suggest that SMAD2/3 KO and DNMT3B KO both impede the naïve to primed transition, which will be especially acute from hypomethylated 2i/LIF cells. The authors should also include WT+DNMT3B sh controls to identify the degree to which this condition also disrupts gene expression in Figure 4E/F.

Reviewer #2

(Remarks to the Author)

The new paper by Zhao and colleagues studies early embryonic development and the role of signaling pathways that belong to the TGF-beta family. Using as model system mouse embryonic stem cells (mESCs) and in vitro cultures that imitate developmental processes, including epiblast generation in culture, the study monitors the functional role of TGF-beta family signaling proteins, the Smads. Emphasis is given to the distinction between Smad2 and Smad3, two specific Smad proteins that become phosphorylated by TGF-beta family receptors from Smad4, a protein that forms complexes with phosphorylated Smad2 and Smad3 and which is known to mediate several physiological processes downstream from the ligands of the TGF-beta family. By employing specific knockout strains of the ESCs, the study shows that epiblast development does not require Smad4, yet it is critically affected by Smad2/3 double knockout. Then, the study monitors genomic loci to which these Smads associate using ChIP-seq methods and identify such genes. Via proteomic screen for proteins that associate with Smad2/3, they identify the DNA methyltransferase Dnmt3b as a new partner of Smad2/3, which mediates specific transcriptional regulation by Smad2/3 during epiblast development. The findings are novel and extend our understanding of the developmental roles that Smad proteins play during mouse embryonic development at least in cultured systems. The interaction and functional role of the Smad2/3-Dnmt3b protein complex is also novel and interesting. Some experiments are well controlled and presented in a clear manner, including statistical analysis, yet some other experiments generate issues as explained below.

Major comments: I have three general and conceptual comments:

1) The study treats Smad2 and Smad3 as a single functional unit. Yet, we know from pioneering studies in mouse development that these two proteins are not redundant and for example lead to drastically different effects when knocked out in mouse embryos. Even the proteomic screen analyzed proteins that interact with Smad2/3 on the basis of an antibody that cannot distinguish between these two different proteins. Thus, one is left with the open question as to whether the new mechanism depends primarily on Smad2 or Smad3.

More specifically, do these 2 proteins interact with equal affinity with Dnmt3b? This raises issues of specificity and interpretation of the data. If we take as example Figure 4, it is obvious that the antibody precipitates both Smad2 and Smad3. But then when a phospho-specific antibody is used, it is also obvious that among the 2 Smad protein bands only the slow migrating band lights up, suggesting that Activin induces only phospho-Smad2 and not phospho-Smad3. Dnmt3b seems to interact with Smad2/3 under all tested conditions, in the absence or presence of Smad4, in the absence or presence of Activin stimulation, and the worse, even in the absence of Smad2/3 in the triple knockout cells (Figure 4C). Thus, overall the mechanistic model of the study is weak and functionally, it is important to distinguish Smad2 from Smad3 as this has already been extensively studied in the past.

2) My second comment has to do with the specificity of the Smad-mediated regulation of genes. This is studied after RNAseq analysis and after ChIPseq analysis, which is very nice. The ChIPseq analysis also combines monitoring of various histone modifications and attempts to correlate Smad binding to histone modifications and also to explain these patterns based on the action of Dnmt3b. I found the latter aspect rather unclear. In order to generate cause and effect relationship between the gene expression patterns and the chromatin-binding patterns, Dnmt3b is silenced by shRNA and various mutant forms of Dnmt3b are analyzed, which is appropriate. The impact of Dnmt3b is analyzed at the level of specific gene expression regulation. I understood that these genes are important biomarkers of the development differentiation process, but data on differentiation per se are missing from the analysis. The study lacks some appropriate controls. For example, the data of Figures 4E, 5E do not show the impact of the shDnmt3b on WT cells. If these controls are the data of Figure S4D, then the gene profiles do not change, so I remain with questions.

3) The above comment relates to my last comment. After reading the paper I could not relate the analysis to the normal, physiological developmental process of differentiation. Are there any circumstances during which embryonic development proceeds in the absence of Smad4? I understand that the Smad4 knockout demonstrates that this protein is not necessary for epiblast formation. But do the genes that are regulated during epiblast formation have Smad4 associated with their chromatin together with Smad2 or Smad3 or not? The model will have broad biological significance if the authors explained the circumstances during which Activin or other TGF-beta family signaling, operate in the absence of Smad4, during early embryogenesis. Maybe I have missed this important control, but I could not identify the Smad4 ChIPseq analysis in the WT condition or at least the ChIP-qPCR analysis of the marker genes in WT cells using the Smad4 antibody.

Minor semantic comment: Figure 6, bottom right shows the Smad2/3 loss condition in which Smad4 forms a trimer and binds to chromatin. To the best of my knowledge, the physiological existence of Smad4 trimers in the absence of R-Smads has not been demonstrated before, thus the stoichiometry of the Smad complex in the Smad2/3 knockout condition remains to be analyzed.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

I would like to thank the authors for their hard work in reorganizing and updating their manuscript, now titled "A stepwise mode of TGFβ-SMAD signaling and DNA methylation regulates naïve-to-primed pluripotency and differentiation." Overall, I find the extra experiments to be very helpful in clarifying their model and connecting the concurrent waves of transcriptional and chromatin regulation that occur during this transition, as well as the dynamics of SMAD2/3 vs SMAD4 signaling as it relates to DNMT3B induction and activity. I also appreciate the thorough efforts to characterize the nature of their biochemical interactions, including control Co-IP experiments and Proximity Ligation Assay. To me, the refined model and presentation of the data is substantially improved and near ready for publication.

I do have some remaining concerns about the panel of ChIP-seq and Flag-tag CUT&Tag assays in Figure 5 and would like to see them sufficiently replicate-powered and cross compared before acceptance. At the moment, it is unclear to me how many times the experiments were conducted, and the supplementary data do not show the degree to which the lower input, epitope-tagged CUT&Tag assays compare genome-wide to the high input ChIP-seq assays. For example, I would like to see data reporting the overlap of annotated peaks, correlation in signal intensity, etc, to showcase that these two technologies report consistently and accurately on DNMT3B genomic occupancy.

I am particularly worried because, by visual inspection, the enrichment intensity for these enzymes over SMAD2/3 peaks appears to be notably different between the epitope-tagged CUT&Tag and the antibody-directed ChIP-seq experiments (see Figures 5a and ED Figure 5a). It has been reported that CUT&Tag can be sensitive to excessive tagmentation under imperfect conditions, leading to artifactual ATAC-seq-like profiles, and this can be especially problematic for factors with broad genomic occupancy such as the DNMT3s (see, for example, PMID: 32913232 for description of these artifacts). To me, all three DNMT3B ChIP/CUT&Tag experiments seem quite variable: the extremely punctate signal within the "Dnmt3b CUT&Tag-2" experiments in Figure 5a don't resemble the DNMT3B ChIP-seq data or "Dnmt3b CUT&Tag-1" data. All told, these differences raise some caution about the authors' interpretations and I ask them to reinvestigate which signal is accurate (this could include the generation of corresponding ATAC-seq data to establish that the signals don't correlate in a manner that would suggest an artifact, as well as further replicate power and/or description/justification of the reaction conditions used).

I'm also surprised that the authors elected to go with a C-terminal Flag tag, as this would be situated proximally to the catalytic domain. Could the authors provide additional data or referenced validation that these epitope-tagged versions are

functional, as well as further supporting data that the integrated transgene is correct (currently we only have ED Fig. 5b for reporting these lines)?

Finally, please refine the model Figure 6 such that DNMT3B does not appear to be interacting with H3K4 methylated (or H3K27ac) chromatin, as this is canonically an inhibitory interaction that blocks DNMT3 binding or de novo activity. I similarly find the conjecture that these interactions are likely to be H3K36me3 related to be a bit overstated (at the very least, the DNMT3 PWWP domains can also interact with H3K36me2) but can understand if the authors would like to keep this interpretation given their rescue experiments with mutant Dnmt3b constructs.

Reviewer #2

(Remarks to the Author)

The revised paper NCOMMS-23-43854A by Zhao and colleagues describes a novel mechanism of the early steps of mouse embryonic stem cell development ex vivo that starts with pluripotent stem cells, proceeds towards the epiblast stage during a naïve to primed transition step and ends with mesendodermal differentiation. Two key factors are studied, the signaling factors Smad2, Smad3 and Smad4 of the TGF-beta family pathways (represented by nodal or activin at early embryonic stages) and the epigenetic factor Dnmt3b that methylates cytosines on CpG-rich DNA sequences. Smad2, Smad3 and Smad4 are known to function as a heterotrimer in a variety of physiological and pathological contexts. The paper presents three tiers of novel mechanisms: a) the early stage naïve to primed transition is driven by Smad2 and Smad3 but Smad4 is dispensable for this biological process. During this stage pluripotency factors like Oct4 and Sox2 are downregulated and Dnmt3b levels are enhanced, the latter contributing to DNA methylation and silencing of the pluripotency genes. b) Mesendodermal differentiation follows, which requires the established Smad2/Smad3/Smad4 transcriptional complex. c) Smad2 and Smad3, but not Smad4, are shown to interact with Dnmt3b, and such complexes position the Smads on hypomethylated gene regions, thus promoting gene expression by the cooperative signal of Smads and Dnmt3b, a protein otherwise linked with gene repression. The positive transcriptional action of Smad2/3-Dnmt3b is also shown to rely on the interaction with histone 3, trimethylated at lysine 36. The paper covers a vast array of experimental methods and well-controlled experiment. The revised paper conveys a clear and novel molecular mechanism that is investigated using a plethora of genome-wide and proteomic technologies and an impressive array of genetic interventions followed by rescue experiments with normal and mutant proteins. The revised paper has enhanced significantly the mechanistic understanding of the Smad2/3-Dnmt3b complex and its chromatin-based association with target genes, on which, this complex acts as a positive regulator of transcription. The paper is succinctly written and also provides a large resource of chromatin-related, protein-related and phenotypic data-sets that can stimulate many new studies in basic stem cell and developmental biology. The authors made an impressive effort to respond to the comments of the reviewers and have delivered an outstanding article.

Based on this latter comment, I recommend the following and final (from my side) minor corrections.

Typographical and minor errors:

1. Line 51: the Smad2-Smad3 is also a heterotrimer. Please correct the term “homotrimer” that implies Smad2-Smad2-Smad2 or Smad3-Smad3-Smad3.
2. Line 61: Smad4 with capital S.
3. Line 106: Smad instead of SMAD as you use the mouse nomenclature throughout your paper.
4. Line 117: Although mESCs depleted from (add the word: from).
5. Line 235: ..., we were curious... or ..., we became curious..., but not ..., we are curious...
6. Line 324: ...for epiblast gene (not genes) expression...
7. Line 336: ...remains constant (not consistent) ...
8. Line 536: It can suppress gene transcription, ...: It refers to DNA methylation, but this is not obvious. Better to be explicit: DNA methylation can suppress gene transcription, ...
9. Line 848: if possible spell the last name correctly: J. Massagué (with an accent on the last e).
10. Lines 860-861: performed the bioinformatic analysis. (not the bioinformatics).
11. The references do not use a single and consistent style when it comes to the journal name.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

I thank the authors again for their work in stabilizing specific claims regarding DNMT3B and TGFb in the naïve to primed transition, including more detailed characterization/validation of their flag-tagged mESC lines and deeper analysis of their CUT&Tag and ChIP-seq data. Ultimately I still find many of the broad conclusions regarding Smad2/3 binding, DNMT3B upregulation and the naïve-to-primed epiblast transition to be compelling and suitable for publication. Conversely, I still believe the CUT&Tag/native ChIP-seq data is quite problematic and in contrast with what is broadly understood about DNMT3 binding and function, particularly during the Naïve to Primed transition where DNMT3B is understood to be performing a global role in remethylating the preimplantation genome. To me, the overlapping genomic occupancy between SMAD2/3 and DNMT3B are likely an artifact of imperfect enrichment techniques and biasing towards nucleosome-depleted euchromatin when IP'ing globally distributed factors. To me, this issue is not easily rectified with additional experiments, but

I think it would be sufficient to publish if the language of this last section and the discussion are more carefully worded.

The authors show in their newly provided figures that their tagmentation-based CUT&Tag data does not overlap with their sonication-based ChIP-seq data (ED Figure 5g) and that this data shows a strong linear relationship with normal ATAC-seq data (ED Figure 5l). Yes, within these scatterplots, the average intensity is higher for the ATAC-seq than it is for DNMT3B CUT&Tag for the majority of peaks, but this is the precise definition of an artifact (a signal that can be explained by a technical detail of the assay, namely the use of Tn5 transposase). Similarly, it is not appropriate for the authors to comment on this assay's improved signal to noise ratio around peaks for regulators that are not expected to generally have peaks, but are rather broadly distributed over the genome (like the DNMT3's).

To me this concern is particularly born out by the correspondingly high normalized signal intensity around the TSS in ED Figure 5J, an enrichment that is very hard to rectify against DNMT3's allosteric inhibition and poor correlation with promoter-enriched H3K4me3 (see previous reports such as Baubec et al, 2015, particularly Figure 3 and ED Figure 2 of that paper that show general depletion around the TSS and no correlation between DNMT3B and the TSS). Yes, ED Figure 5J shows that there is gene-body enrichment in the CUT&Tag vs ATAC-seq, which appears to be true signal. In contrast I would presume the authors agree that the TSS enrichment is an "artifact" compatible with basal tagmentation of euchromatin. Given that the SMAD binding sites are likely to be similarly sensitive to Tn5, how are we supposed to distinguish this non-canonical and somewhat exotic model of cooperative DNMT3B binding from an artifact?

I do not think it's fair to ask the authors to rectify this challenging aspect of their manuscript with more experiments, particularly after several rounds of review where the explicit biochemical mechanism has been a sticking point. Instead, I ask that they soften the language in the last section of the manuscript, including language that insists that the CUT&Tag data is better based upon signal to noise ratio, and to comment on which features of ED Figure 5 the authors believe to be true or not based upon the literature.

Then, I would ask them to add a paragraph to the discussion about how it remains unclear precisely how DNMT3B may be molecularly interacting with the Nodal signaling pathway during mESC differentiation, including an acknowledgement that it is difficult to distinguish whether or not the broader ChIP-seq data (and findings from papers such as Baubec et al.) and more punctate CUT&Tag data disagree because of the improved nature of CUT&Tag or because it is not an ideal assay to study regulators with global functions. Please also highlight in this paragraph the challenging kinetics and cellular heterogeneity inherent to stem cell-based differentiation assays, especially given the rapid and possibly overlapping transitions between naïve to primed and primed to mesendoderm. The conclusions that are most appropriate are that SMAD2/3 induces DNMT3B, DNMT3B & SMADs appear to biochemically interact during a window where they are both active and highly expressed, and this window precedes co-binding by SMAD4. If the authors appropriately highlight the current caveats as room for future investigation, they are welcome to maintain Figure 6 as a "potential" or "proposed" model.

Please also more generally refer to H3K36 methylation (or me2/3) throughout the manuscript rather than H3K36me3, which is generally associated with transcriptional elongation and not enriched at enhancers.

credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

General response to reviewers

We highly appreciate all the reviewers' positive feedback, including their finding that our manuscript is "an interesting manuscript with several notable findings", "The findings are novel and extend our understanding of the developmental roles that Smad proteins play during mouse embryonic development", and it "make a series of interesting connections between several major genomic and transcriptional/regulatory dynamics that accompany early mouse embryogenesis". Moreover, all the reviewers' insightful suggestions have greatly enhanced our understanding and interpretation of the data in this work.

We have now addressed all the comments with the substantial addition of new experimental data and bioinformatic analyses. In the following pages, please find our point-by-point response to all specific comments of the reviewers. The reviewers' comments are in **Black**, and our responses to them are in **Blue**. Please note that the reviewers' comments address some common issues, resulting in similar queries. If questions overlap, we have answered each one individually rather than referring the reviewer to another response. This has led to some redundancy among the answers below, but our intent is to make each response as clear as possible. For some specific experiments, we have included the related figures and legends below our response for the reviewers' quick reference. Otherwise, please refer to the cited figures in the revised manuscript. Moreover, we have highlighted the changes described below in the revised manuscript in **Blue** to further facilitate review.

Finally, we sincerely thank all the reviewers for their valuable and insightful comments and suggestions that helped us make substantial improvements to the revised manuscript.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

In their manuscript “Dnmt3b mediates TGFB-Smad2/3 action during epiblast priming from mouse embryonic stem cells independently of Smad4,” Zhao, Yu, Shi, Ma et al. make a series of interesting connections between several major genomic and transcriptional/regulatory dynamics that accompany early mouse embryogenesis. Beginning with generation of Smad2/3 and Smad4 KO mouse embryonic stem cell lines, the authors identify several waves of Smad binding and activity, including a clear role for Smad2/3 in enabling the naïve to primed transition from conventional mouse ESCs to mouse Epiblast-like cells (EpiLCs), which is followed by an auxiliary/secondary role of Smad4 in inducing mesendodermal regulators such as Eomes and Wnt3. Then, the authors use these cells to identify context-specific interaction partners via IP-Mass Spec, including an interaction between the DNA methyltransferase DNMT3B and SMAD2/3 within Smad4 KO cells, an interaction that is not as apparent in WT EBs. The authors then develop and test a model that postulates an intermediate role for this interaction either prior to or instead of canonical SMAD2/3-4-based gene regulation.

Overall, this is an interesting manuscript with several notable findings related to the stepwise progression of early postimplantation development, including developmental signal-transduction (Nodal-based Smad signaling), naïve-to-primed changes in chromatin state (hypo to hypermethylated genomes), and differentiation (Smad4-facilitated primitive streak induction). However, these very nice connections within the early portion of the manuscript are somewhat unnaturally merged with a biochemical model that is much harder to address and, to me, relies on weaker data. Overall, I feel that rectifying the strengths of these earlier figures with the biochemical information in Figures 4 and 5 will be quite challenging, but offer some comments on how to possibly address them in greater detail below.

We greatly appreciate and are encouraged that this reviewer found our work to be of interesting with several notable findings. Following the reviewer's constructive comments and insightful suggestions, we have adjusted the paper title, refined the text and flow, revised the working model, conducted the recommended and additional experiments to strengthen our work. We believe these changes have led to a substantially improved manuscript. Below, please find our point-by-point response to the reviewer's comments.

To me, the biochemical interactions between SMAD2/3 and DNMT3B could very well be indirect and simply enriched within Smad4 KO EBs because this line produces a different combination of less differentiated cell types than either WT or Smad2/3 KO EBs. A streamlined model would be that Smad2/3 signaling is dominant in the naïve to primed transition (evident in Figures 2 and 3), induces DNMT3B, and this interaction

is captured by nature of their shared nuclear localization and high expression. Then, Smad4 participates in downstream differentiation of primed state cells into mesodermal or endodermal cell types, which involves a more specific interaction with SMAD2/3 and promotes downstream cell types where DNMT3B is less actively engaged in global genome remethylation. In other words, the apparent interaction between SMAD2/3 and DNMT3B could very well be because these different KO lines enrich for different cell types along the pluripotent to differentiated continuum, with Smad4 KO EBs consisting of more pre-differentiated epiblast-like and early primitive streak cells. Similarly, Dnmt3B KD and OE experiments could phenocopy various results by altering the methylation status of primed state cells, which will rely on this modification for viability.

Ultimately, disentangling such fast and heterogeneous dynamics using bulk epigenomic and biochemical assays would be very tricky to impossible, and I do not feel that the very interesting aspects of this manuscript should be over-ridden by these technical limitations. As an alternative, I would ask that the authors consider making these other potential explanations clear in the text as well as rewording the title, abstract and model figure (Figure 6) to prioritize the strongest aspects of their data, namely a stepwise series of events that implicates a transcriptional role of Smad2/3 in supporting DNMT3B induction, enrichment for nuclear localization and potential interactions between Smad2/3 and DNMT3B, followed by formation of Smad2/3-4 heterotrimers that appear to support primitive streak induction.

We deeply thank the reviewer for raising this important point and for helping us refine our understanding of the critical dynamics during the early post-implantation development. As per the reviewer's suggestion, we have implemented the following changes in the revised manuscript:

1) We have changed the title to “A stepwise mode of TGF β -SMAD signaling and DNA methylation regulates naïve-to-primed pluripotency and differentiation”, and we welcome any further suggestions.

2) In the abstract, main text, and discussion sections, we have fully discussed the stepwise paradigm in which Smad2/3 regulate the lineage priming and differentiation of mouse embryonic stem cells (mESCs) by collaborating with different effectors. For example, we highlighted the model in the abstract: "During the naïve-to-primed transition, Smad2/3 upregulate DNA methyltransferase 3b (Dnmt3b), which establishes the proper DNA methylation patterns and, in turn, enables Smad2/3 binding to the hypomethylated centers of promoters and enhancers of epiblast marker genes. Consequently, in the absence of Smad2/3, Smad4 alone cannot initiate epiblast-specific gene transcription. When primed epiblast cells begin to differentiate, Dnmt3b becomes less actively engaged in global genome methylation, and Smad4 takes over the baton in this relay race, forming a complex with Smad2/3 to support mesendoderm induction. Thus, mESCs lacking Smad4 can undergo the priming

process but struggle with the downstream differentiation. "

3) We also updated the model in Figure 6, which is attached below. First, Smad2/3 play a dominant role in orchestrating the transition from the naïve mESC to the primed epiblast stage by inducing Dnmt3b. on one hand, Dnmt3b establishes DNA methylation patterns and ensures the timely downregulation of pluripotent factors like Sox2 by methylating DNA at DMRs. On the other hand, Dnmt3b interacts with Smad2/3, facilitating their binding to the hypomethylated centers of promoters and enhancers of epiblast genes. Second, Smad2/3-Smad4 complex play a key role during mesendoderm differentiation from the primed EpiLCs, in which Dnmt3b is downregulated and becomes less actively involved. The Smad2/3-Smad4 complex binds to the promoters and enhancers of mesendoderm genes, with the prebound pioneer factor Foxh1^{4, 5}.

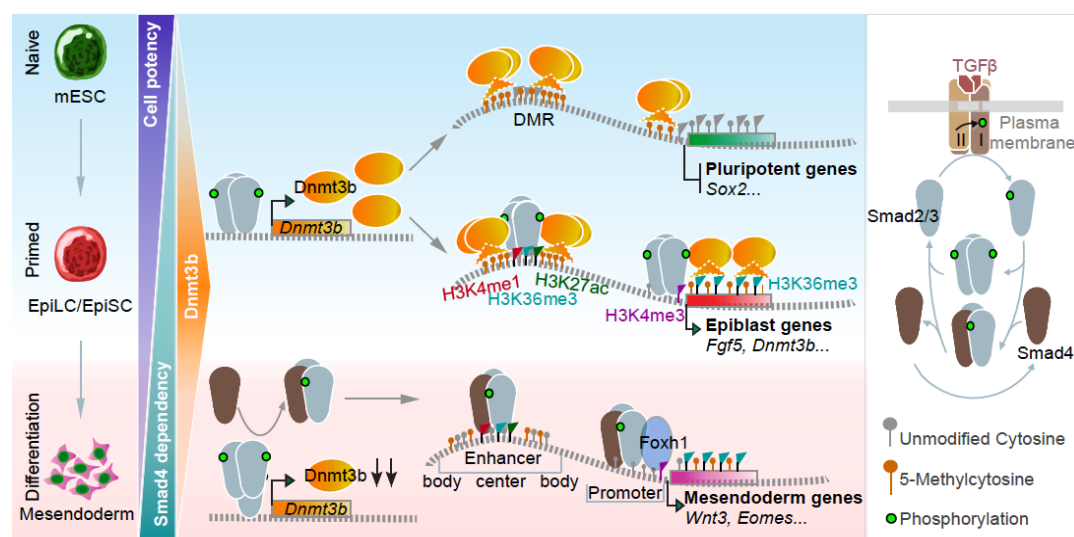


Figure 6. A two-step model centering on Smad2/3 describes the transition from naïve to primed epiblast and subsequent differentiation.

The figures themselves could also be further streamlined with substantial data being brought into the supplementary data – Figures 1, 2 and 3 ultimately address very similar points and could be reduced to two figures total. The key details to highlight are the distinct transcriptional dynamics between either Smad2/3 vs Smad4 KO, validation that Smad2/3 is sufficient for EpiLC-induction, and ChIP-seq data demonstrating that Smad2/3 binding precedes Smad4 binding and includes Dnmt3b locus.

We thank the reviewer for this important suggestion, which has improved the flow and coherence of our narrative. In this revised manuscript, Figures 1, 2 and 3 from the previous version have been reorganized into Revised Figures 1 and 2.

Revised Figure 1 focuses on the distinct transcriptional dynamics between Smad2/3 knockout (S2/3DKO) or Smad4 knockout (KO) and emphasizes the differential requirements for Smad2/3 and Smad4 during mESC primed transition and differentiation. To strengthen this conclusion, we have juxtaposed a new heatmap

showing primitive streak or mesendoderm genes in WT, S2/3DKO and S4KO (Revised Figure 1I) with the one for epiblast genes (Revised Figure 1H). Revised Figure 2 mainly demonstrates that Smad2/3 binding to epiblast genes precedes Smad4 binding and predominantly drives the EpiLC transition (Revised Figure 2).

The rest of the data have been moved into the Revised Supplementary Figure 1 and 2. Additionally, we used a two-step mesendoderm differentiation model (mESC-EpiLC-Mesendoderm)¹ to further confirm whether both Smad2/3 and Smad4 are required for mesendoderm differentiation from EpiLCs. Consistent with our findings in the EB models, the loss of either Smad2/3 or Smad4 inhibits the induction of mesendoderm genes such as *Wnt3*, *Eomes* and *Gsc* (Revised Supplemental Figure 2D). CUT&Tag experiments showed that the binding of Smad2/3 to these mesendoderm genes drastically decreased with the loss of Smad4, indicating the crucial role of the canonical Smad2/3-Smad4 complex in promoting mesendoderm differentiation from EpiLCs (Revised Supplemental Figure 2E).

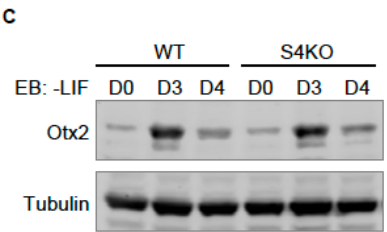
Figure 4 and 5 seem especially problematic because they assume the interactions between SMAD2/3 and DNMT3B are direct, exist as an alternative interaction than SMAD4, and speak to the combined actions of these factors at chromatin. Alternatively, I would suggest that Smad4 KO EBs are less (or differently) differentiated than WT, have higher genomic DNMT3 activity, and that DNMT3B co-immunoprecipitates with SMAD2/3 under stimulating conditions because they are both nuclear. To substantiate their claim, I feel the authors need to establish that the SMAD2/3-DNMT3B interaction is specific, generalizable (not an artifact of cell state composition) and occurs at chromatin:

We fully agree that it is crucial to determine whether the interaction between Smad2/3 and Dnmt3b is direct, or simply enriched in the expanded population of pre-differentiated epiblast-like cells due to Smad4 loss (S4KO). We think that both mechanisms could be true, contributing to the cooperation between signal transduction and chromatin state. mESCs lacking Smad4 can undergo the priming process but struggle with downstream differentiation (Figure 1I-J, Figure 2A-B, and Figure S2D-E). Therefore, in S4KO EBs, both free Smad2/3 and highly expressed Dnmt3b are enriched in the nucleus, facilitating their interaction (Figure 3). Moreover, we confirmed this interaction in mESC E14Tg2a and human embryonic kidney 293T (HEK293T), and by pull-down assay using GST-Dnmt3b and His-Smad2 or 3 induced in *E. coli* (Revised Figure 3C-E and 3G). In addition, we visualized the interaction of Dnmt3b and Smad2/3 in WT, but more prominently in S4KO EBs, utilizing the in situ proximity ligation assay (PLA) (Revised Figure 3F). This robust technique allows the detection of two proteins in close vicinity (distance < 40 nm)⁶. As a control, no positive signal was observed in S2/3DKO EBs with AC addition (Revised Figure 3F). Detailed experiments and discussions on this matter can be found in the subsequent sections.

- To address specificity, I would ask that the authors consider IP'ing other factors that

become induced or localized to the nucleus during the Naïve to Primed transition, such as pERK or Otx2, to highlight that DNMT3B is not just very abundant during this transition and interacting non-specifically with other nuclear factors.

We thank the reviewer for this important suggestion. We chose Otx2 for comparison because it is known to be highly expressed in the nuclei of epiblast or EpiSC^{7,8}, and it was upregulated during EB differentiation (Revised Figure S3C).



Supplemental Figure 3C: Western blot displayed the protein levels of Otx2 in WT and S4KO during EB differentiation (D0, D3, D4). Tubulin serves as a loading control.

Dnmt3b was pulled down by Smad2//3, while Otx2 was not immunoprecipitated with Smad2/3 in chromatin fractions of WT and S4KO EBs (Revised Figure 3D).

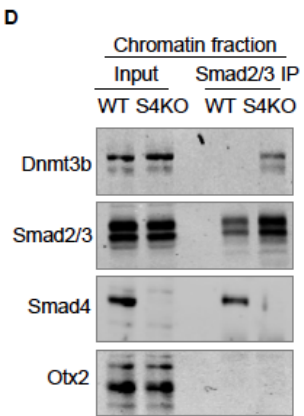
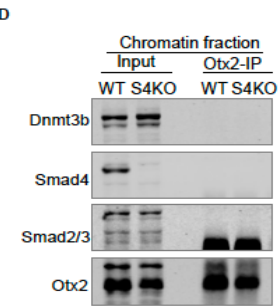


Figure 3D: Smad2/3 immunoprecipitation in the chromatin fraction of D3 EBs in WT and S4KO with AC treatment were analyzed by western blot using Dnmt3b, Smad2/3, Smad4 and Otx2 antibodies.

Meanwhile, the specific Otx2 antibody pulled down itself but failed to precipitate Dnmt3b, Smad2/3 or Smad4 from all the chromatin samples (Revised Figure S3D).



Supplemental Figure 3D: Western blot analysis of Dnmt3b, Smad2/3, Smad4 and Otx2 in chromatin fraction of WT and S4KO EBs after immunoprecipitation using anti-Otx2 antibody.

- To address generalizability, I would ask that the authors consider validating these interactions in an ectopic cell system, even something as straightforward as HEK293 cells.

We thank the reviewer for the great suggestion. Following the reviewer's advice, we utilized the following two systems:

1) human embryonic kidney 293T (HEK293T) cells were utilized, overexpressing HA-S2, HA-S3, or/and Flag-tagged Dnmt3b (Flag-Dnmt3b). Co-IP assays revealed that Flag-Dnmt3b efficiently precipitated either HA-S2 or HA-S3 individually, as well as their co-expression (Revised Figure 3F).

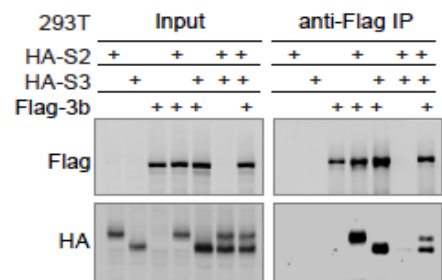


Figure 3F: Overexpression of HA tagged Smad2 (HA-S2) or Smad3 (HA-S3), or/and Flag tagged Dnmt3b (Flag-3b) in 293T cells were shown in the figure. Western blot of Flag and HA were detected after anti-Flag immunoprecipitation.

2) In addition, Smad2 and Smad3 were individually expressed as 6XHistidine fusion proteins (His-S2 and His-S3), while GST-tagged Dnmt3b (GST-3b) was induced in *E. coli*. These proteins were subsequently employed in pull-down assays. GST-3b, but not GST or GST-GFP, was found to interact with His-S2 or His-S3 (Revised Figure 3G-H).

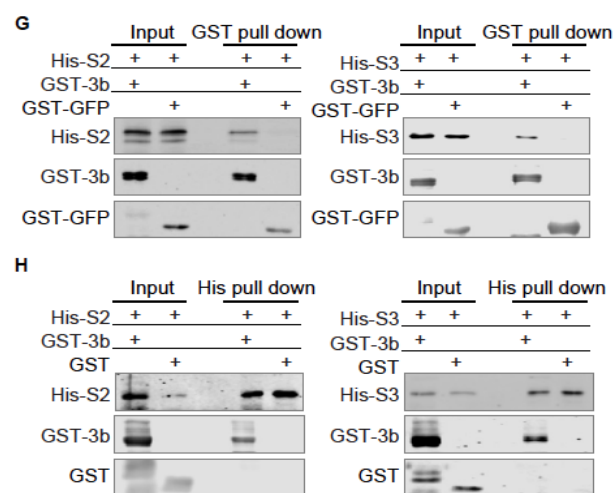


Figure 3: G. GST pull-down assays using bacterially expressed GST-Dnmt3b or GST-GFP, and 6xHistidine-tagged Smad2 (His-S2, left) or Smad3 (His-S3, right). Elution proteins were analyzed using anti-GST and anti-His antibodies. H. His pull-down assays using bacterially expressed GST-Dnmt3b or GST, and 6xHistidine-tagged Smad2 (His-S2, left) or Smad3 (His-S3, right). Elution proteins were analyzed using anti-GST and anti-His antibodies.

- To address their activity or interaction at chromatin, I would ask that the authors consider fractionating their cell models to demonstrate that the SMAD2/3-DNMT3B interaction is specific and at chromatin, rather than generic interactions within the cytoplasm or nucleoplasm. They could also consider altering the stringency of their IP buffers.

We thank the reviewer for the insightful suggestions. As the reviewer suggested, we explore whether Smad2/3-Dnmt3b interaction occurs at chromatin. Both WT and S4KO EBs with AC addition were fractionated to get chromatin. As a control, Smad2/3 immunoprecipitated Smad4 in WT. Dnmt3b, but not otx2 was successfully pulled down with Smad2/3 from chromatin fraction of S4KO (Revised Figure 3D).

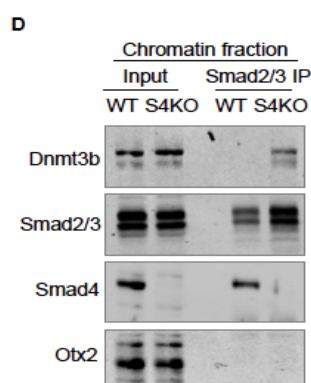


Figure 3D: Smad2/3 co-IP experiments were done with the chromatin fraction of D3 EBs in WT and S4KO with AC treatment, which were analyzed by western blot using Dnmt3b, Smad2/3, Smad4 and Otx2 antibodies.

Furthermore, in the Revised Figure 5, CUT&Tag and ChIP-seq analyses confirmed that both Dnmt3b and Smad2/3 converge on epiblast genes, and the binding pattern of Dnmt3b is highly concordant with Smad2/3 binding genome-wide. (Revised Figure 5A-B and S5A). These data were discussed in detail below.

To me, Figure 5 does not highlight much beyond what is already known about DNMT3B recruitment or expression dynamics and doesn't show much convincing evidence that DNMT3B or SMAD2/3 are colocalized. If anything, Figure 5E fairly confidently establishes that they are enriched at different genomic features. Most comments about differential activity/behavior at the Fgf5 promoter is based on very weak signal of a single region. If the authors are motivated to showcase the degree to which these factors co-occupy similar regions of the genome, they should consider a genome-wide approach (DNMT3B ChIP-seq in their WT and KO cell types). Odds are fairly good though that these will largely confirm that these factors go to different genomic regions. Should the authors attempt a more genome-wide investigation, I would also ask that they include DNMT3A as it can be more clearly associated with other genomic features beyond gene bodies.

We thank the reviewer for raising this important point. To explore this, we conducted a

series of Dnmt3b CUT&Tag and ChIP-seq experiments during EB differentiation (Revised Figure 5). For comparison, we also performed these experiments for Dnmt3a. Besides using a specific Dnmt3a or Dnmt3b antibody, we also employed anti-Flag antibody to detect the Dnmt3a/3b binding pattern in Dnmt3a/3b-Flag cells, in which a 3XFlag epitope tag was inserted into the C-terminus of endogenous Dnmt3a (3a-Flag) and Dnmt3b (3b-Flag) by CRISPR/Cas9 technology, respectively (Revised Figure S5B).

1) CUT&Tag and ChIP-seq analyses confirmed that both Dnmt3b and Smad2/3 bind to the epiblast genes, including *Fgf5* promoter and enhancers (Figure 5A). The binding pattern of Dnmt3b was highly concordant with Smad2/3 binding genome-wide (Figure 5A-B and S5A). The binding peaks of Dnmt3b were strongly disrupted by Dnmt3b-shRNA. This disruption demonstrates the specificity of these peaks (Revised Figure 5A and S5A).

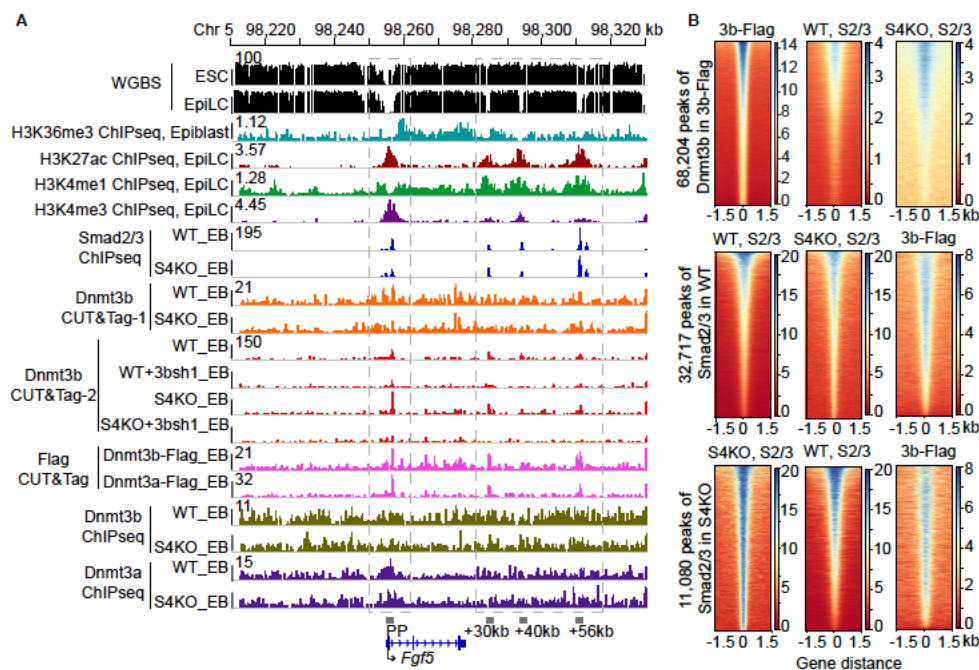
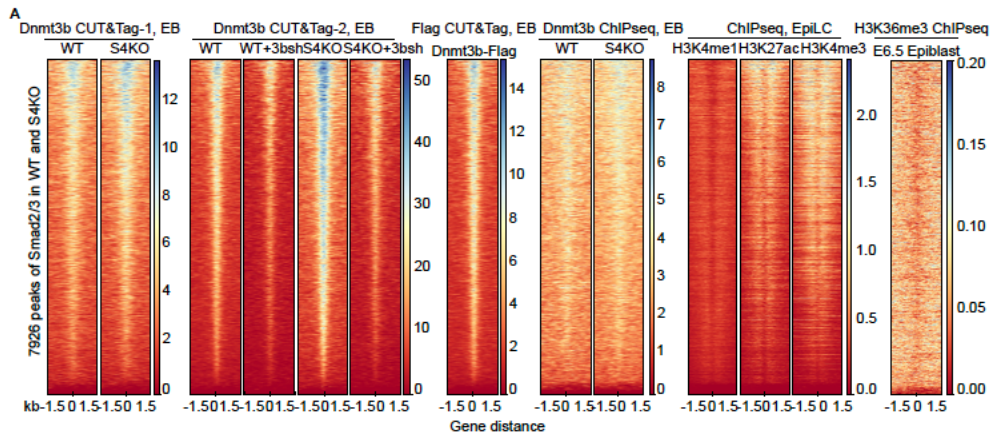


Figure 5. Dnmt3b mediates Smad2/3 binding to epiblast genes independently of Smad4.

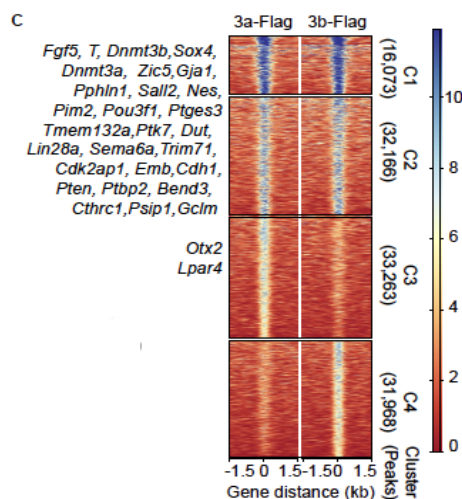
A. IGV view of WGBS, ChIP-seq and CUT&Tag profiles surrounding *Fgf5* gene locus. WGBS as well as ChIP-seq of H3K27ac, H3K4me1 and H3K4me3 were performed at ESC or EpiLC stages¹, H3K36me3 ChIP-seq were performed in E6.5 epiblast of mice². Smad2/3 ChIP-seq were done in EBs of WT³ and S4KO (this study), Dnmt3b CUT&Tag of EBs were carried out in WT, S4KO, WT+Dnmt3b-shRNA1 (WT+3bsh1) and S4KO+Dnmt3b-shRNA1 (S4KO+3bsh1), Flag CUT&Tag experiments were performed in Dnmt3a-Flag or Dnmt3b-Flag cells; Dnmt3a or Dnmt3b ChIP-seq of WT and S4KO EBs were shown at the bottom.

B. Heatmap of tag densities for Flag CUT&Tag in Dnmt3b-Flag, and Smad2/3 ChIP-seq in WT or S4KO, were shown within -1.5/+1.5kb genomic regions surrounding the centers of 68,204 Dnmt3b binding sites in Dnmt3b-Flag EBs, 32,717 or 11,080 Smad2/3 binding sites in WT or S4KO EBs, respectively.



Supplemental Figure 5A: Heatmaps of CUT&Tag and ChIP-seq tag densities were within -1.5/+1.5kb genomic regions surrounding the centers of these 7,926 Smad4-independent Smad2/3 binding peaks.

2) We also compared the genome-wide pattern of Dnmt3a and Dnmt3b by CUT&Tag and ChIP-seq analysis. Although both Dnmt3a and Dnmt3b colocalized on the PP and enhancers of *Fgf5*, the binding signals in gene body regions disclosed with anti-Flag in Dnmt3b-Flag cells were less detectable in Dnmt3a-Flag cells (Revised Figure 5A). This discrepancy in binding preference was also evident in the ChIP-seq data for Dnmt3b and Dnmt3a (Revised Figure 5A). Furthermore, as shown on the heatmap, all the binding sites of Dnmt3a-Flag and Dnmt3b-Flag were analyzed and divided into four clusters (C1-C4). Epiblast genes were highly enriched in C1 and C2, while in C3 and C4, Dnmt3a and Dnmt3b showed different binding abilities (Revised Figure S5C). Collectively, these findings indicate the overlapping yet distinct functions of Dnmt3a and Dnmt3b.

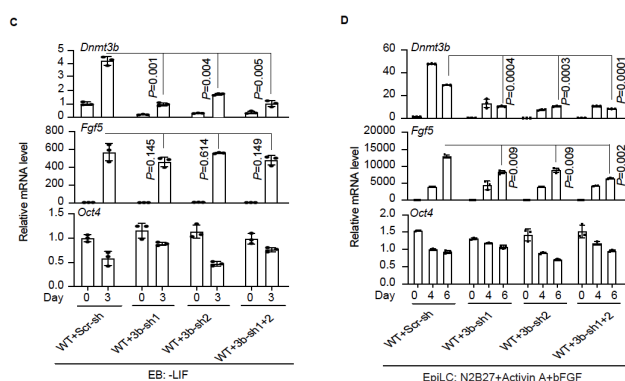


Supplemental Figure 5C: Heatmap of Flag CUT&Tag densities for 3a-Flag and 3b-Flag in EBs. Representative epiblast maker genes in C1-C2 and C3 were annotated.

Furthermore, the inability to rescue delayed developmental defects with anything other than the full, catalytically active DNMT3B would suggest that SMAD23 KOs and DNMT3B KO both impede the naïve to primed transition, which will be especially acute

from hypomethylated 2i/LIF cells. The authors should also include WT+DNMT3B sh controls to identify the degree to which this condition also disrupts gene expression in Figure 4E/F.

As the reviewer suggested, we knocked down Dnmt3b in WT using two different shRNAs or a combination of both (WT+Dnmt3b-sh1, 2, or 1+2) and performed EB differentiation or EpiLC transition assays under the same conditions as in the original Figures 4E and 4F. Scramble-shRNA (Scr-sh) was used as a control. When Dnmt3b was repressed in WT, the epiblast gene *Fgf5* was mildly affected in the more heterogeneous EB models but had a greater effect in the more homogeneous EpiLCs (Supplemental figure 4C and 4D).



Supplemental Figure 4C and 4D: C. The mRNA levels of *Dnmt3b*, *Fgf5* and *Oct4* when Dnmt3b was silenced with shRNA in D0 or D3 WT EBs. The *P* value of each sample was compared to WT and calculated by two-way ANOVA test. D. Expression of *Dnmt3b*, *Fgf5* and *Oct4* during EpiLC transition (D0, D4, D6) after Dnmt3b knockdown in WT were measured by qRT-PCR. The *P* value of each sample was compared to WT and calculated by two-way ANOVA test.

Reviewer #2 (Remarks to the Author):

The new paper by Zhao and colleagues studies early embryonic development and the role of signaling pathways that belong to the TGF-beta family. Using as model system mouse embryonic stem cells (mESCs) and in vitro cultures that imitate developmental processes, including epiblast generation in culture, the study monitors the functional role of TGF-beta family signaling proteins, the Smads. Emphasis is given to the distinction between Smad2 and Smad3, two specific Smad proteins that become phosphorylated by TGF-beta family receptors from Smad4, a protein that forms complexes with phosphorylated Smad2 and Smad3 and which is known to mediate several physiological processes downstream from the ligands of the TGF-beta family. By employing specific knockout strains of the ESCs, the study shows that epiblast development does not require Smad4, yet it is critically affected by Smad2/3 double knockout. Then, the study monitors genomic loci to which these Smads associate using ChIP-seq methods and identify such genes. Via proteomic screen for proteins that associate with Smad2/3, they identify the DNA methyltransferase Dnmt3b as a

new partner of Smad2/3, which mediates specific transcriptional regulation by Smad2/3 during epiblast development. The findings are novel and extend our understanding of the developmental roles that Smad proteins play during mouse embryonic development at least in cultured systems. The interaction and functional role of the Smad2/3-Dnmt3b protein complex is also novel and interesting. Some experiments are well controlled and presented in a clear manner, including statistical analysis, yet some other experiments generate issues as explained below.

We greatly appreciate the reviewer for the positive feedback and for highlighting our novel findings. Following the reviewer's suggestions, we have conducted a series of additional experiments and provide more in-depth bioinformatic analysis to strengthen our conclusions. Please see our point-by-point response below:

Major comments: I have three general and conceptual comments:

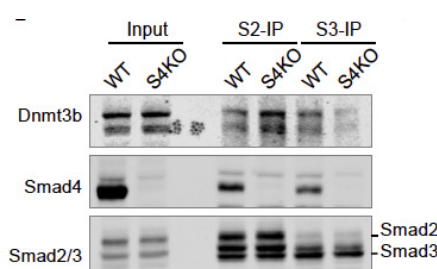
1) The study treats Smad2 and Smad3 as a single functional unit. Yet, we know from pioneering studies in mouse development that these two proteins are not redundant and for example lead to drastically different effects when knocked out in mouse embryos. Even the proteomic screen analyzed proteins that interact with Smad2/3 on the basis of an antibody that cannot distinguish between these two different proteins. Thus, one is left with the open question as to whether the new mechanism depends primarily on Smad2 or Smad3.

More specifically, do these 2 proteins interact with equal affinity with Dnmt3b? This raises issues of specificity and interpretation of the data. If we take as example Figure 4, it is obvious that the antibody precipitates both Smad2 and Smad3. But then when a phospho-specific antibody is used, it is also obvious that among the 2 Smad protein bands only the slow migrating band lights up, suggesting that Activin induces only phospho-Smad2 and not phospho-Smad3. Dnmt3b seems to interact with Smad2/3 under all tested conditions, in the absence or presence of Smad4, in the absence or presence of Activin stimulation, and the worse, even in the absence of Smad2/3 in the triple knockout cells (Figure 4C). Thus, overall the mechanistic model of the study is weak and functionally, it is important to distinguish Smad2 from Smad3 as this has already been extensively studied in the past.

We thank the reviewer for this important suggestion. We totally agree that Smad2 and Smad3 are not identical. Although Smad2 and Smad3 share a considerable amino acid sequence similarity, they could exhibit distinct functionalities in various biological contexts^{5, 9-11}. It has been reported that Smad2 surpasses Smad3 in abundance in mESCs and early mouse embryos^{5, 12}. In mESCs, a Smad2:Smad3 abundance ratio is around 6:1. During EB development, this ratio shifts to 4:1 by day 3⁵. Consistently, both Smad2 and its phosphorylated form, pSmad2, were more effectively captured by anti-Smad2/3 antibody than their Smad3 counterparts (Figure 3B, upper bands vs. lower bands). Please note that the lower bands may include Smad3 and Smad2 β due to their similar molecular weight⁵. To study the individual roles of Smad2 and Smad3, we conducted the following experiments to investigate whether the interaction between

Dnmt3b and Smad2/3 depends primarily on Smad2 or Smad3.

1) A specific Smad2 or Smad3 antibody was used for co-IP experiment in EBs stimulated with AC (Revised Figure S3E). Smad4 was pulled down successfully by Smad2 or Smad3 in WT, but Dnmt3b exhibited a much higher efficiency of immunoprecipitation when bound to Smad2, especially in the S4KO EBs (Revised Figure S3F). Thus, Smad2 is the primary protein responsible for pulling down Dnmt3b using the anti-Smad2/3 antibody in mESCs, especially considering the higher expression level of Smad2 compared to Smad3. But at this point, we cannot totally rule out the role of Smad3 in the interaction due to potential antibody efficiency issues and expression discrepancies between Smad2 and Smad3.



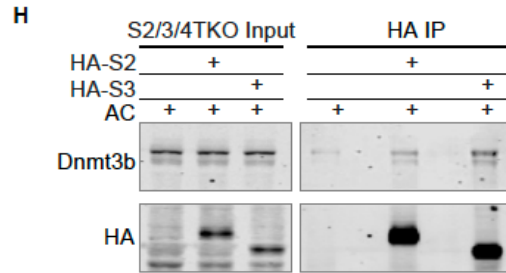
Supplemental Figure 3F: Dnmt3b, Smad2/3 and Smad4 were analyzed by western blot after Smad2 and Smad3 immunoprecipitation in WT and S4KO EBs.

2) In addition, co-IP experiments conducted in Smad2 knockout (S2KO) or Smad3 knockout (S3KO) cells⁵ using an anti-Smad2/3 antibody showed that individual Smad2 or Smad3 could pull down Dnmt3b. Their IP efficiencies were also related to their expression levels (Figure S3G).



Supplemental Figure 3G: Smad2/3 immunoprecipitation in S2KO and S3KO D3 EBs under AC addition were analyzed by western blot using Dnmt3b, Smad2/3 and Smad4 antibodies.

3) Considering the variance in Smad2 and Smad3 abundance in mESC, we sought to standardize their expression levels. Therefore, we overexpressed Smad2 and Smad3 to comparable levels in S2/3/4TKO cells. Remarkably, under these conditions, both HA-S2 and HA-S3 demonstrated efficient interaction with endogenous Dnmt3b under AC stimulation (Figure S3H).



Supplemental Figure 3H: Western blot analysis of Dnmt3b and HA expression for the anti-HA immunoprecipitation in Smad2/3/4TKO cells, in which HA tagged Smad 2 or 3 (HA-S2 and HA-S3) were overexpressed with AC stimulation.

4) Furthermore, human embryonic kidney 293T (HEK293T) cells were utilized, overexpressing HA-S2, HA-S3, or/and Flag-tagged Dnmt3b (Flag-Dnmt3b). Co-IP assays revealed that Flag-Dnmt3b efficiently precipitated either HA-S2 or HA-S3 individually, as well as their co-expression (Revised Figure 3F).

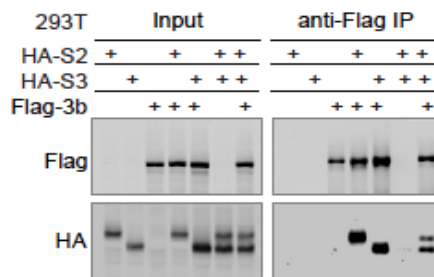


Figure 3F: Overexpression of HA tagged Smad2 (HA-S2) or Smad3 (HA-S3), or/and Flag tagged Dnmt3b (Flag-3b) in 293T cells were shown in the figure. Western blot of Flag and HA were detected after anti-Flag immunoprecipitation.

5) Finally, Smad2 and Smad3 were individually expressed as 6XHis fusion proteins (His-S2 and His-S3), while GST-tagged Dnmt3b (GST-3b) was induced in *E. coli*.

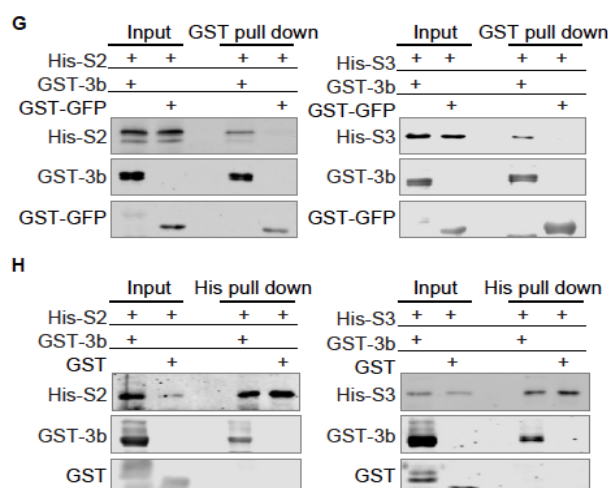


Figure 3: G. GST pull-down assays using bacterially expressed GST-Dnmt3b or GST-GFP, and 6xHistidine-tagged Smad2 (His-S2, left) or Smad3 (His-S3, right). Elution proteins were analyzed using anti-GST and anti-His antibodies. H. His pull-down assays using bacterially expressed GST-Dnmt3b or GST, and 6xHistidine-tagged Smad2 (His-S2, left) or Smad3 (His-S3, right). Elution proteins were analyzed using anti-GST and anti-His antibodies.

These proteins were subsequently employed in pull-down assays. GST-3b, but not GST or GST-GFP, was found to interact with His-S2 or His-S3 (Revised Figure 3G-H).

In summary, in mESC, Smad2 is more abundant and interacts more strongly with Dnmt3b. But Smad2 and Smad3 could potentially interact with Dnmt3b at similar levels if their expression levels are similar.

Moreover, as the reviewer pointed out that the previous Figure 4C displayed immunoprecipitation (IP) signals even in the absence of Smad2/3 in the triple knockout cells. We apologize for the lack for clarity. The IP signal observed in the S2/3/4TKO cells was background noise due to the longer exposure time used. We have replaced this figure with one captured using a shorter exposure time to avoid any confusion (Revised Figure 3C).

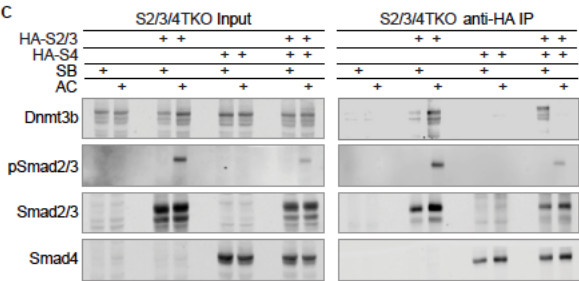
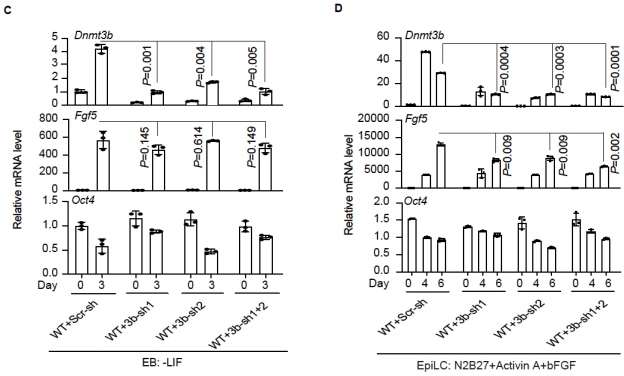


Figure 3C: Western blot analysis of Dnmt3b, pSmad2/3, Smad2/3 and Smad4 for the anti-HA immunoprecipitation in Smad2/3/4TKO cells in which HA tagged Smad2/3 (HA-S2/3) or/and Smad4 (HA-S4) were overexpressed as indicated in the figure. Cells were collected after SB or AC treatment.

2) My second comment has to do with the specificity of the Smad-mediated regulation of genes. This is studied after RNAseq analysis and after ChIPseq analysis, which is very nice. The ChIPseq analysis also combines monitoring of various histone modifications and attempts to correlate Smad binding to histone modifications and also to explain these patterns based on the action of Dnmt3b. I found the latter aspect rather unclear. In order to generate cause and effect relationship between the gene expression patterns and the chromatin-binding patterns, Dnmt3b is silenced by shRNA and various mutant forms of Dnmt3b are analyzed, which is appropriate. The impact of Dnmt3b is analyzed at the level of specific gene expression regulation. I understood that these genes are important biomarkers of the development differentiation process, but data on differentiation per se are missing from the analysis. The study lacks some appropriate controls. For example, the data of Figures 4E, 5E do not show the impact of the shDnmt3b on WT cells. If these controls are the data of Figure S4D, then the gene profiles do not change, so I remain with questions.

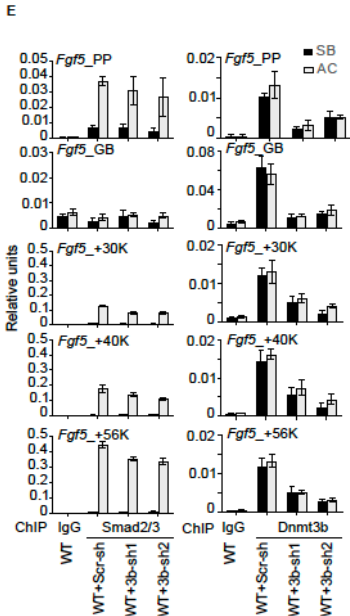
We appreciate the reviewer for raising this important point. We conducted the following experiments and bioinformatics analyses to determine how Dnmt3b affect Smad2/3-mediated epiblast formation.

1) As the reviewer suggested, we knocked down *Dnmt3b* in WT using two different shRNAs or a combination of both (WT+*Dnmt3b*-sh1, 2, or 1+2) and performed EB differentiation or EpiLC transition assays under the same conditions as in the original Figures 4E and 4F. Scramble-shRNA (Scr-sh) was used as a control. When *Dnmt3b* was repressed in WT, the epiblast gene *Fgf5* was mildly affected in the more heterogeneous EB models but had a greater effect in the more homogeneous EpiLCs (Revised supplemental Figure 4C and 4D).



Supplemental Figure 4C and 4D: C. The mRNA levels of *Dnmt3b*, *Fgf5* and *Oct4* when *Dnmt3b* was silenced with shRNA in D0 or D3 WT EBs. The *P* value of each sample was compared to WT and calculated by paired t-test. D. Expression of *Dnmt3b*, *Fgf5* and *Oct4* during EpiLC transition (D0, D4, D6) after *Dnmt3b* knockdown in WT were measured by qRT-PCR. The *P* value of each sample was compared to WT and calculated by paired t-test.

2) When *Dnmt3b* was interfered with using the same shRNAs in WT EBs, *Smad2/3* binding to *Fgf5* was mildly affected in WT EBs as shown by ChIP-qPCR (Revised Figure S5E).



Supplemental Figure 5E: ChIP-qPCR analysis of *Smad2/3* and *Dnmt3b* binding to *Fgf5* promoter (Fgf5_PP), gene body (Fgf5_GB) and enhancers (Fgf5_+30k, Fgf5_+40k, and Fgf5_+56k) of D3 EBs treated with SB or AC in WT, WT+ Scr-sh, WT+3b-Sh1 or WT+3b-sh2 cell lines. ChIP with anti-IgG serves as a negative control. Error bars represent SD.

In summary, Dnmt3b is required for the priming process, and this requirement is particularly amplified in the context of S4KO. Considering the competitive binding of Dnmt3b and Smad4 to Smad2/3, and the more pre-differentiated epiblast-like status in S4KO compared to WT EBs, Dnmt3b has a higher chance to interact with Smad2/3 upon loss of Smad4. In addition to its role in DNA methylation, we postulate that Dnmt3b enables Smad2/3 binding to epiblast genes, which is required for transcription activation. This further increases the dependence of Smad2/3 on Dnmt3b. It is difficult to disentangle such fast and heterogeneous dynamics during mESC differentiation, but S4KO cells provide a platform to achieve this. But of course, we cannot rule out other mechanisms, which requires further study in the future.

3) To strengthen the conclusion that Dnmt3b enables Smad2/3 binding to epiblast genes, we compared the genome-wide binding pattern of Dnmt3b and Smad2/3 by CUT&Tag and ChIPseq in WT and S4KO EBs (Revised Figure 5). Interestingly, both

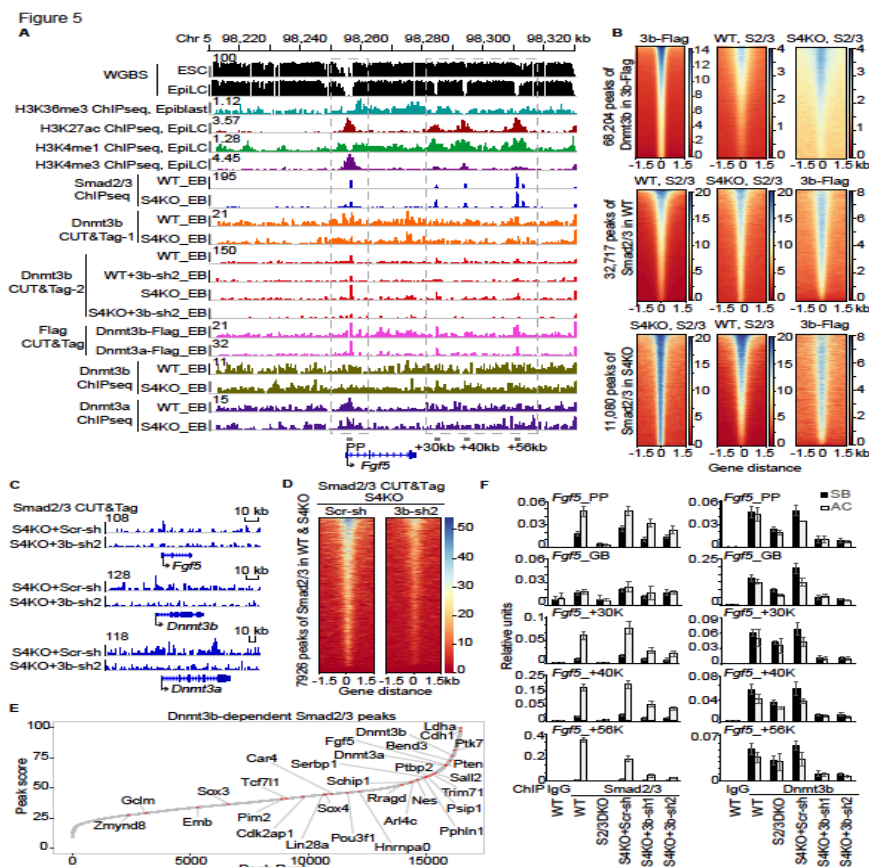
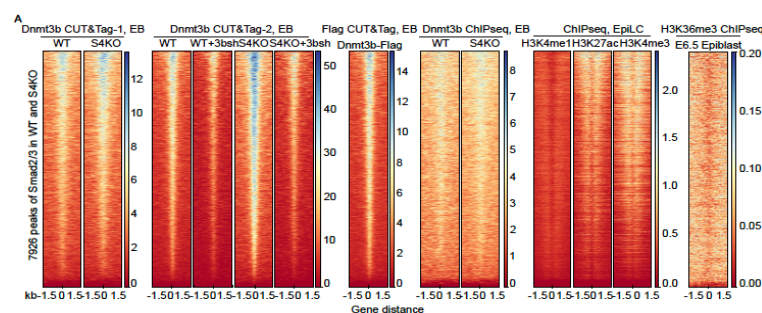


Figure 5. Dnmt3b mediates Smad2/3 binding to epiblast genes independently of Smad4.

A. IGV view of WGBS, ChIP-seq and CUT&Tag profiles surrounding *Fgf5* gene locus. B. Heatmap of tag densities for Flag CUT&Tag in Dnmt3b-Flag, and Smad2/3 ChIP-seq in WT or S4KO, were shown within -1.5/+1.5kb genomic regions surrounding the centers of 68,204 Dnmt3b binding sites in Dnmt3b-Flag EBs, 32,717 or 11,080 Smad2/3 binding sites in WT or S4KO EBs, respectively. C. Gene track view of the *Fgf5*, *Dnmt3a*, and *Dnmt3b* loci from Smad2/3 CUT&Tag in EBs of S4KO and S4KO+3b-sh2. D. Heatmap of Smad2/3 CUT&Tag signals in S4KO and S4KO+3bsh2 surrounding 7,926 Smad4-independent Smad2/3 peaks. E. Peak score plot for the most significant Dnmt3b-dependent Smad2/3 binding sites (fold change > 2.0, $P < 0.01$). Epiblast marker genes adjacent to the peaks are indicated. F. ChIP-qPCR analysis of Smad2/3 and Dnmt3b binding to *Fgf5* promoter (*Fgf5*_PP), gene body (*Fgf5*_GB) and enhancers (*Fgf5*_+30k, *Fgf5*_+40k, and *Fgf5*_+56k) of D3 EBs treated with SB or AC in WT, S2/3KO, S4KO+ Scr-shRNA, S4KO+3b-Sh1 or S4KO+3b-sh2 cell lines. ChIP with anti-IgG serves as a negative control. Error bars represent SD.

Dnmt3b and Smad2/3 bind to the epiblast genes, including *Fgf5* promoter and enhancers, which were further confirmed by ChIP-qPCR (Revised Figure 5A-B and 5F). In addition, the binding pattern of Dnmt3b was highly concordant with Smad2/3 binding genome-wide (Revised Figure 5A-B and S5A). These binding peaks of Dnmt3b were strongly disrupted by Dnmt3b-shRNA, demonstrating the specificity of these peaks (Revised Figure 5A and S5A). Furthermore, Smad2/3 binding peaks in S4KO EBs were strongly diminished in by Dnmt3b-shRNA (Revised Figure 5C-D). These Dnmt3b-dependent Smad2/3 binding peaks showed a high enrichment of epiblast specification genes (Revised Figure 5C and 5E). Indeed, gene track views by IGV and ChIP-qPCR assays confirmed that Dnmt3b depletion inhibited Smad2/3 binding to a subset of epiblast genes, including *Fgf5*, *Dnmt3b*, and *Dnmt3a* (Revised Figure 5C and 5F). Taken together, Dnmt3b potentially facilitates Smad2/3 binding to epiblast genes including *Fgf5*.



Supplemental Figure 5A: Heatmaps of CUT&Tag and ChIP-seq tag densities were within -1.5/+1.5kb genomic regions surrounding the centers of these 7,926 Smad4-independent Smad2/3 binding peaks.

3) The above comment relates to my last comment. After reading the paper I could not relate the analysis to the normal, physiological developmental process of differentiation. Are there any circumstances during which embryonic development proceeds in the absence of Smad4? I understand that the Smad4 knockout demonstrates that this protein is not necessary for epiblast formation. But do the genes that are regulated during epiblast formation have Smad4 associated with their chromatin together with Smad2 or Smad3 or not? The model will have broad biological significance if the authors explained the circumstances during which Activin or other TGF-beta family signaling, operate in the absence of Smad4, during early embryogenesis. Maybe I have missed this important control, but I could not identify the Smad4 ChIP-seq analysis in the WT condition or at least the ChIP-qPCR analysis of the marker genes in WT cells using the Smad4 antibody.

We thank the reviewer for raising this important questions.

1) We apologize for the lack of clarity. In previous version, Smad4 ChIP-seq in WT was shown in Figure 3 and Figure S3A. In current version, Smad4 ChIP-seq was analyzed in the Revised Figure 2C-E. Although the binding pattern is very similar, the overall binding peaks of Smad4 ChIP-seq (13,313 peaks in WT) were not as pronounced as those for Smad2/3 (32,717 peaks in WT), possibly due to differential antibody affinities

(Revised Figure 2C-D). The disappearance of 4,474 Smad2/3 peaks upon Smad4 loss, which were co-bound with Smad4 in WT EBs, strongly suggests a specific requirement for Smad4 at these loci (Revised Figure 2C). Intriguingly, 7,926 Smad2/3 binding peaks remained in the absence of Smad4. Among these, 3,787 Smad2/3 peaks were generally co-occupied by both Smad2/3 and Smad4 (Revised Figure 2C). We annotated genes adjacent to these Smad4-independent 7,926 peaks, in which epiblast marker genes such as *Fgf5*, *Dnmt3a/3b*, *T*, *Sox3/4*, *Pou3f1* and many others were enriched (Revised Figure 2C). The Gene track view showed that in ESCs or SB-treated EBs, there was little to no signal of Smad2/3 or Smad4 on *Fgf5*, *Dnmt3a*, *Dnmt3b*, and *T* genes. In response to the addition of AC to EBs, these typical epiblast genes were co-occupied by Smad2/3 and Smad4 at promoter and distal enhancer regions (Figure 2D-E). If Smad4 was removed, Smad2/3 still retained its binding to these epiblast genes (Figure 2E). In summary, these data suggest that Smad4 participates in signaling transduction, but it is not necessary for Smad2/3 binding. Smad2/3 binding may precede Smad4 binding on epiblast genes, including the *Dnmt3b* locus.

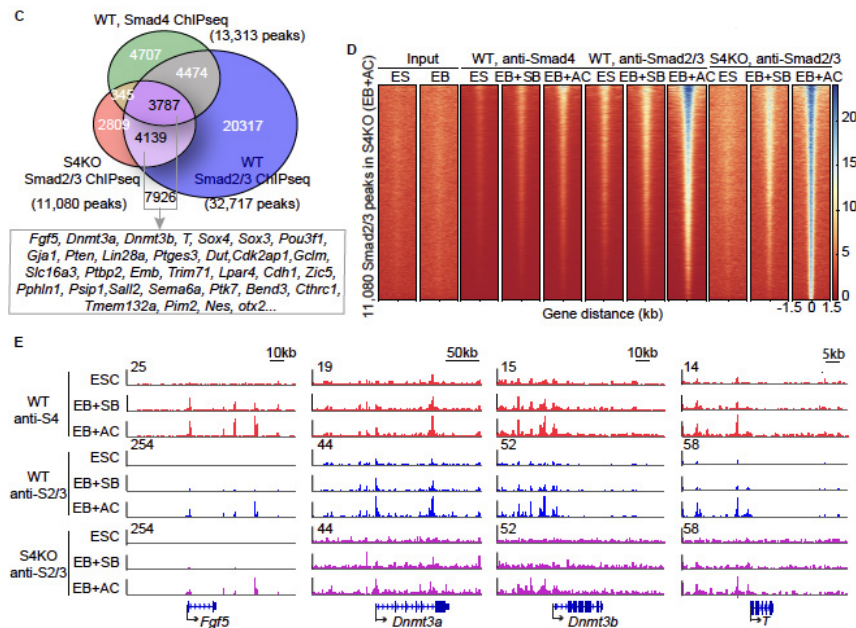


Figure 2C-E: C. Venn diagram indicated the overlapping peaks among Smad4 ChIP-seq in WT, Smad2/3 ChIP-seq in WT and Smad2/3 ChIP-seq in S4KO. Genes adjacent to 7,926 peaks were annotated, in which primed epiblast genes were marked below. D. Heatmap of ChIP-seq tag densities for Smad2/3 and Smad4 were within -1.5/+1.5kb genomic regions surrounding the centers of 11,080 high-confidence Smad2/3 binding sites in S4KO D3 EBs with AC treatment. SB, SB431542, a selective inhibitor of TGF β R; AC, Activin A, an available ligand for Nodal/Activin receptors. E. Gene track view of the *Fgf5*, *Dnmt3a*, *Dnmt3b* and *T* loci. Smad2/3 and Smad4 ChIP-seq were performed in SB- or AC-treated D3 EBs and in D0 non-treated mESCs. Tag densities were normalized to reads per kilobase per million mapped reads (RPKM). The gene structures from RefSeq are schematically represented at the bottom.

2) We deeply thank two reviewers for their constructive comments and have updated the working model in the revised manuscript. We introduce a stepwise paradigm in

which Smad2 and Smad3 (Smad2/3) regulate the lineage priming and differentiation of mESCs during early embryogenesis by collaboration with different effectors. During the naïve-to-primed transition, Smad2/3 upregulate DNA methyltransferase 3b (Dnmt3b), which establishes the proper DNA methylation patterns and, in turn, enables Smad2/3 binding to the hypomethylated centers of promoters and enhancers of epiblast marker genes. Consequently, in the absence of Smad2/3, Smad4 alone cannot initiate epiblast-specific gene transcription. When primed epiblast cells begin to differentiate, Dnmt3b becomes less actively engaged in global genome methylation, and Smad4 takes over the baton in this relay race, forming a complex with Smad2/3 to support mesendoderm induction. Thus, mESCs lacking Smad4 can undergo the priming process but struggle with the downstream differentiation.

Our two-step model is also consistent with the *in vivo* data¹³⁻¹⁵. Conditional knockout of Smad4 in the epiblast using the Cre-LoxP system does not prevent the development from the epiblast to the gastrulation initiation, but shows focal defects in the primitive streak, suggesting the limited role of Smad4 during the priming process¹⁵. As gastrulation proceeds, Smad4-deficient embryos fail to form derivatives of the anterior primitive streak, including definitive endoderm, sharing many phenotypic similarities with the downregulation of the Nodal/Smad2/3/Foxh1 pathway¹³⁻¹⁵. Thus, Smad4 has a limited role during the priming process but is required for primitive streak/mesendoderm differentiation.

Minor semantic comment: Figure 6, bottom right shows the Smad2/3 loss condition in which Smad4 forms a trimer and binds to chromatin. To the best of my knowledge, the physiological existence of Smad4 trimers in the absence of R-Smads has not been demonstrated before, thus the stoichiometry of the Smad complex in the Smad2/3 knockout condition remains to be analyzed.

We apologize for the inaccuracy and thank the reviewer for the important comments. We have revised the model as shown in the Revised Figure 6.

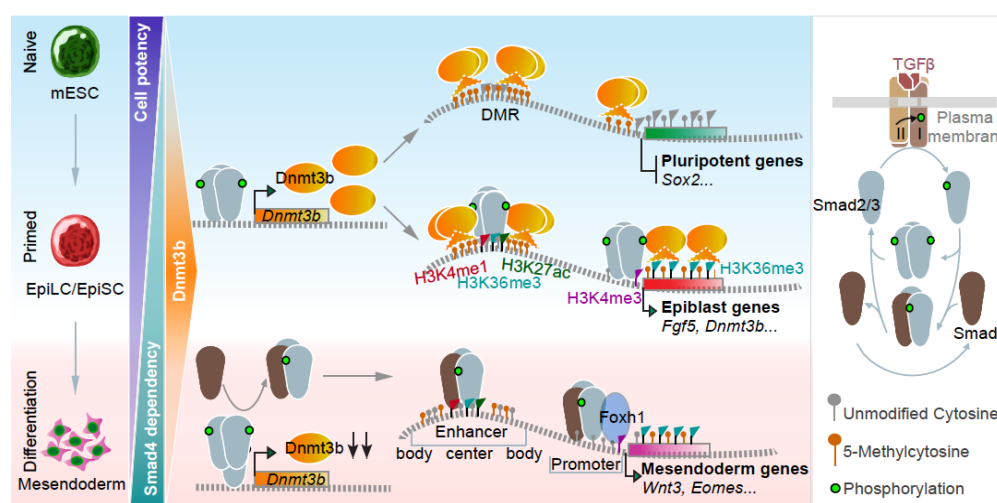


Figure 6. A two-step model describes the transition from naïve to primed epiblast and subsequent differentiation.

Refence

1. Lauria, A., Meng, G., Proserpio, V., Rapelli, S., Maldotti, M., Polignano, I.L., Anselmi, F., Incarnato, D., Krepelova, A., Donna, D., Levra Levron, C., Donati, G., Molineris, I., Neri, F. & Oliviero, S. Dnmt3b supports meso-endoderm differentiation from mouse embryonic stem cells. *Nat Commun* 2023. 14, 367.
2. Hanna, C.W., Perez-Palacios, R., Gahurova, L., Schubert, M., Krueger, F., Biggins, L., Andrews, S., Colome-Tatche, M., Bourc'his, D., Dean, W. & Kelsey, G. Endogenous retroviral insertions drive non-canonical imprinting in extra-embryonic tissues. *Genome Biol* 2019. 20, 225.
3. Wang, Q., Zou, Y., Nowotschin, S., Kim, S.Y., Li, Q.V., Soh, C.L., Su, J., Zhang, C., Shu, W., Xi, Q., Huangfu, D., Hadjantonakis, A.K. & Massague, J. The p53 family coordinates wnt and nodal inputs in mesendodermal differentiation of embryonic stem cells. *Cell stem cell* 2017. 20, 70-86.
4. Charney, R.M., Forouzmamand, E., Cho, J.S., Cheung, J., Paraiso, K.D., Yasuoka, Y., Takahashi, S., Taira, M., Blitz, I.L., Xie, X. & Cho, K.W. Foxh1 occupies cis-regulatory modules prior to dynamic transcription factor interactions controlling the mesendoderm gene program. *Dev Cell* 2017. 40, 595-607 e594.
5. Aragon, E., Wang, Q., Zou, Y., Morgani, S.M., Ruiz, L., Kaczmarek, Z., Su, J., Torner, C., Tian, L., Hu, J., Shu, W., Agrawal, S., Gomes, T., Marquez, J.A., Hadjantonakis, A.K., Macias, M.J. & Massague, J. Structural basis for distinct roles of smad2 and smad3 in foxh1 pioneer-directed tgfbeta signaling. *Genes Dev* 2019. 33, 1506-1524.
6. Soderberg, O., Gullberg, M., Jarvius, M., Ridderstrale, K., Leuchowius, K.J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.G. & Landegren, U. Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat Methods* 2006. 3, 995-1000.
7. Iwafuchi-Doi, M., Matsuda, K., Murakami, K., Niwa, H., Tesar, P.J., Aruga, J., Matsuo, I. & Kondoh, H. Transcriptional regulatory networks in epiblast cells and during anterior neural plate development as modeled in epiblast stem cells. *Development* 2012. 139, 3926-3937.
8. Acampora, D., Di Giovannantonio, L.G. & Simeone, A. Otx2 is an intrinsic determinant of the embryonic stem cell state and is required for transition to a stable epiblast stem cell condition. *Development* 2013. 140, 43-55.
9. Nomura, M. & Li, E. Smad2 role in mesoderm formation, left-right patterning and craniofacial development. *Nature* 1998. 393, 786-790.
10. Zhu, Y., Richardson, J.A., Parada, L.F. & Graff, J.M. Smad3 mutant mice develop metastatic colorectal cancer. *Cell* 1998. 94, 703-714.
11. Jayaraman, L. & Massague, J. Distinct oligomeric states of smad proteins in the transforming growth factor-beta pathway. *J Biol Chem* 2000. 275, 40710-40717.
12. Peng, G., Suo, S., Chen, J., Chen, W., Liu, C., Yu, F., Wang, R., Chen, S., Sun, N., Cui, G., Song, L., Tam, P.P., Han, J.D. & Jing, N. Spatial transcriptome for the molecular annotation of lineage fates and cell identity in mid-gastrula mouse embryo. *Dev Cell* 2016. 36, 681-697.
13. Dunn, N.R., Vincent, S.D., Oxburgh, L., Robertson, E.J. & Bikoff, E.K. Combinatorial activities of smad2 and smad3 regulate mesoderm formation and patterning in the mouse embryo. *Development* 2004. 131, 1717-1728.
14. Sirard, C., de la Pompa, J.L., Elia, A., Itie, A., Mirtsos, C., Cheung, A., Hahn, S., Wakeham, A., Schwartz, L., Kern, S.E., Rossant, J. & Mak, T.W. The tumor suppressor gene smad4/dpc4 is required for gastrulation and later for anterior development of the mouse embryo. *Genes Dev* 1998. 12, 107-119.
15. Chu, G.C., Dunn, N.R., Anderson, D.C., Oxburgh, L. & Robertson, E.J. Differential requirements for smad4 in tgfbeta-dependent patterning of the early mouse embryo. *Development* 2004. 131, 3501-3512.

General response to reviewers

We greatly appreciate the reviewers' positive feedback and recognition of the improvements made to the paper.

We have now addressed all the remaining comments by incorporating new experimental data and additional bioinformatic analyses. In the following pages, you will find our point-by-point responses to each of the reviewers' comments. The reviewers' comments are presented in **Black**, with our responses in **Blue**. For specific experiments, we have included the corresponding figures and legends below our responses for the reviewers' convenience. Otherwise, please refer to the cited figures in the revised manuscript. Additionally, changes described in the responses have been highlighted in **Blue** in the revised manuscript to further facilitate the review process.

Once again, we sincerely thank all the reviewers for their valuable comments and suggestions, which have helped us make substantial improvements to the revised manuscript.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

I would like to thank the authors for their hard work in reorganizing and updating their manuscript, now titled ““A stepwise mode of TGF β -SMAD signaling and DNA methylation regulates naïve-to-primed pluripotency and differentiation.” Overall, I find the extra experiments to be very helpful in clarifying their model and connecting the concurrent waves of transcriptional and chromatin regulation that occur during this transition, as well as the dynamics of SMAD2/3 vs SMAD4 signaling as it relates to DNMT3B induction and activity. I also appreciate the thorough efforts to characterize the nature of their biochemical interactions, including control Co-IP experiments and Proximity Ligation Assay. To me, the refined model and presentation of the data is substantially improved and near ready for publication.

I do have some remaining concerns about the panel of ChIP-seq and Flag-tag CUT&Tag assays in Figure 5 and would like to see them sufficiently replicate-powered and cross compared before acceptance. At the moment, it is unclear to me how many times the experiments were conducted, and the supplementary data do not show the degree to which the lower input, epitope-tagged CUT&Tag assays compare genome-wide to the high input ChIP-seq assays. For example, I would like to see data reporting the overlap of annotated peaks, correlation in signal intensity, etc, to showcase that these two technologies report consistently and accurately on DNMT3B genomic occupancy.

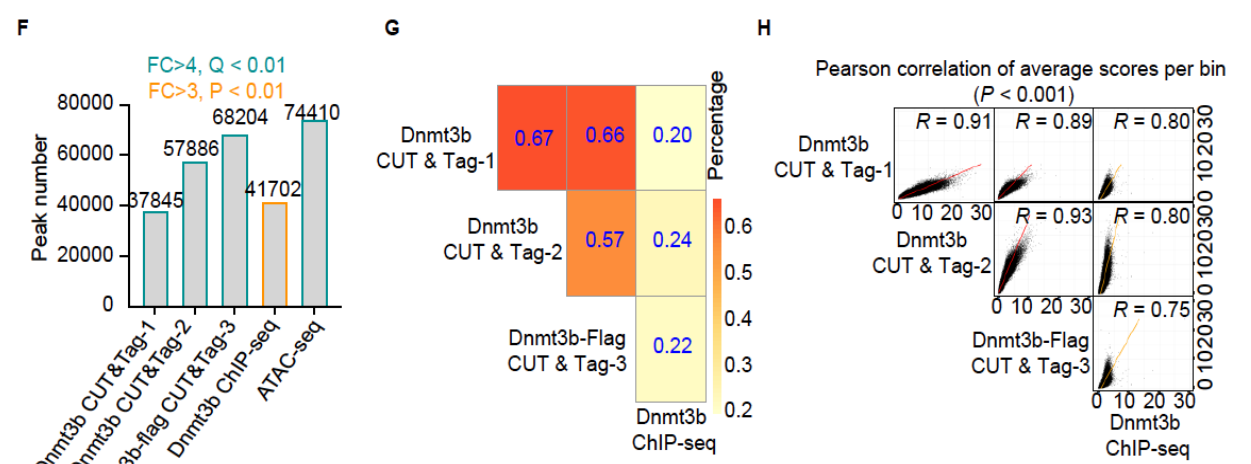
We greatly appreciate and are encouraged by the reviewer's recognition that our work is substantially improved and nearly ready for publication. We also thank the reviewer for raising this important questions and apologize for the lack of clarity.

1) We have conducted three CUT&Tag assays and one ChIP-seq assay. CUT&Tag-1, CUT&Tag-2, and ChIP-seq, all were performed with an anti-Dnmt3b antibody, and CUT&Tag-3 used an anti-Flag antibody in Dnmt3b-Flag cells, in which a 3XFlag epitope tag was inserted into the C-terminus of endogenous Dnmt3b. Notably, CUT&Tag-2 included an E. coli genome spike-in for calibration, which minimized errors related to low input and sequencing biases, improving comparability between samples.

2) As the reviewer suggested, we compared Dnmt3b binding profiles across the four assays. In the three CUT&Tag assays, we identified 37,845, 57,886, and 68,204 Dnmt3b peaks (fold change ≥ 4 , $Q < 0.01$), with 57%-67% of annotated peaks overlapping, demonstrating high reproducibility among them (Revised Supplementary Figure 5F-G). In contrast, the ChIP-seq assay, due to its lower peak-calling performance, identified 41,702 Dnmt3b peaks even with a more lenient fold change cutoff of 3 and a P-value of 0.01 (Revised Supplementary Figure 5F).

3) Additionally, only about 20% of annotated peaks overlapped between the ChIP-seq and CUT&Tag assays (Revised Supplementary Figure 5G). Furthermore, we divided the genome into 1 kb bins and calculated the Pearson correlation (R) of signal intensity between each pair of assays. The R values among the three CUT&Tag datasets ranged from 0.89 to 0.93 (Revised Supplementary Figure 5H), demonstrating high reproducibility of Dnmt3b CUT&Tag biological

replicates, consistent with a published report¹. In contrast, the R values between Dnmt3b CUT&Tag and ChIP-seq assays were slightly lower, ranging from 0.75 to 0.80 (Revised Supplementary Figure 5H).



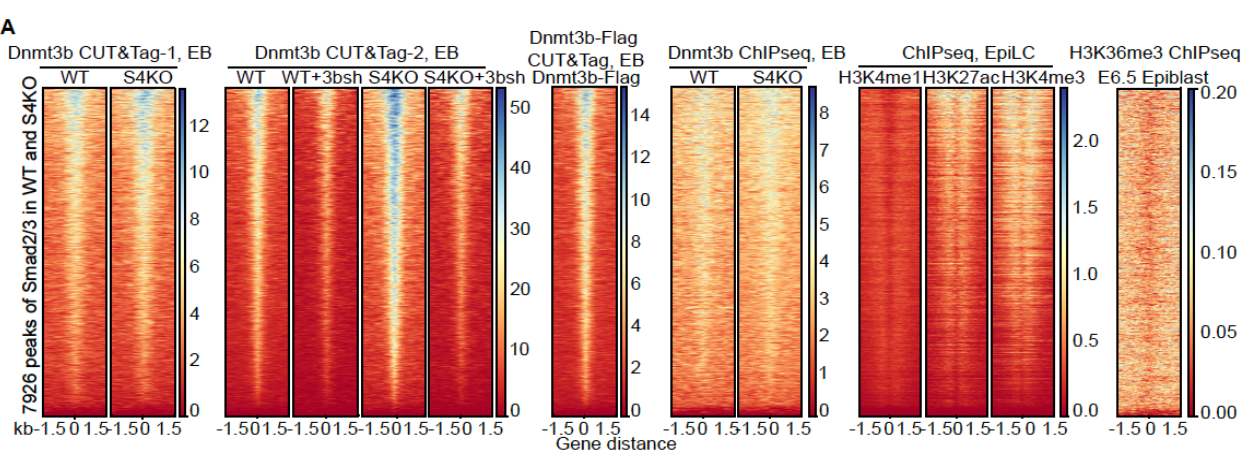
Supplementary Figure 5. The genome-wide binding profiles of Dnmt3b.

F. Bar plots showing the number of Dnmt3b binding peaks identified by MACS2 in three CUT & Tag assays and one ChIP-seq assay, as well as ATAC-seq peaks in D3 EBs. CUT & Tag peaks and ATAC-seq peaks were called with a fold change cutoff of 4 and a Q-value of 0.01(Cyan), while ChIP-seq peaks used a fold change cutoff of 3 and a P-value of 0.01(yellow).

G. Heatmap displays the overlap percentages of Dnmt3b binding peaks between each pair of experiments as indicated.

H. Scatterplots show the Pearson correlation (R) of signal intensity for Dnmt3b binding profiles across the whole genome, which were divided into 1 kb bins, between each pair of assays. An F-test was performed. R values were shown as indicated.

4) Additionally, at 7,926 Smad2/3 binding sites independent of Smad4, all four assays show consistent Dnmt3b binding patterns, which were significantly reduced in Dnmt3b-shRNA samples (Revised Supplementary Figure 6A), further confirming the accuracy of Dnmt3b binding.



Supplementary Figure 6. Dnmt3b and Smad2/3 converge on epiblast genes across the genome.

A. Heatmaps of CUT&Tag and ChIP-seq tag densities were within -1.5/+1.5kb genomic regions surrounding the centers of these 7,926 Smad4-independent Smad2/3 binding peaks.

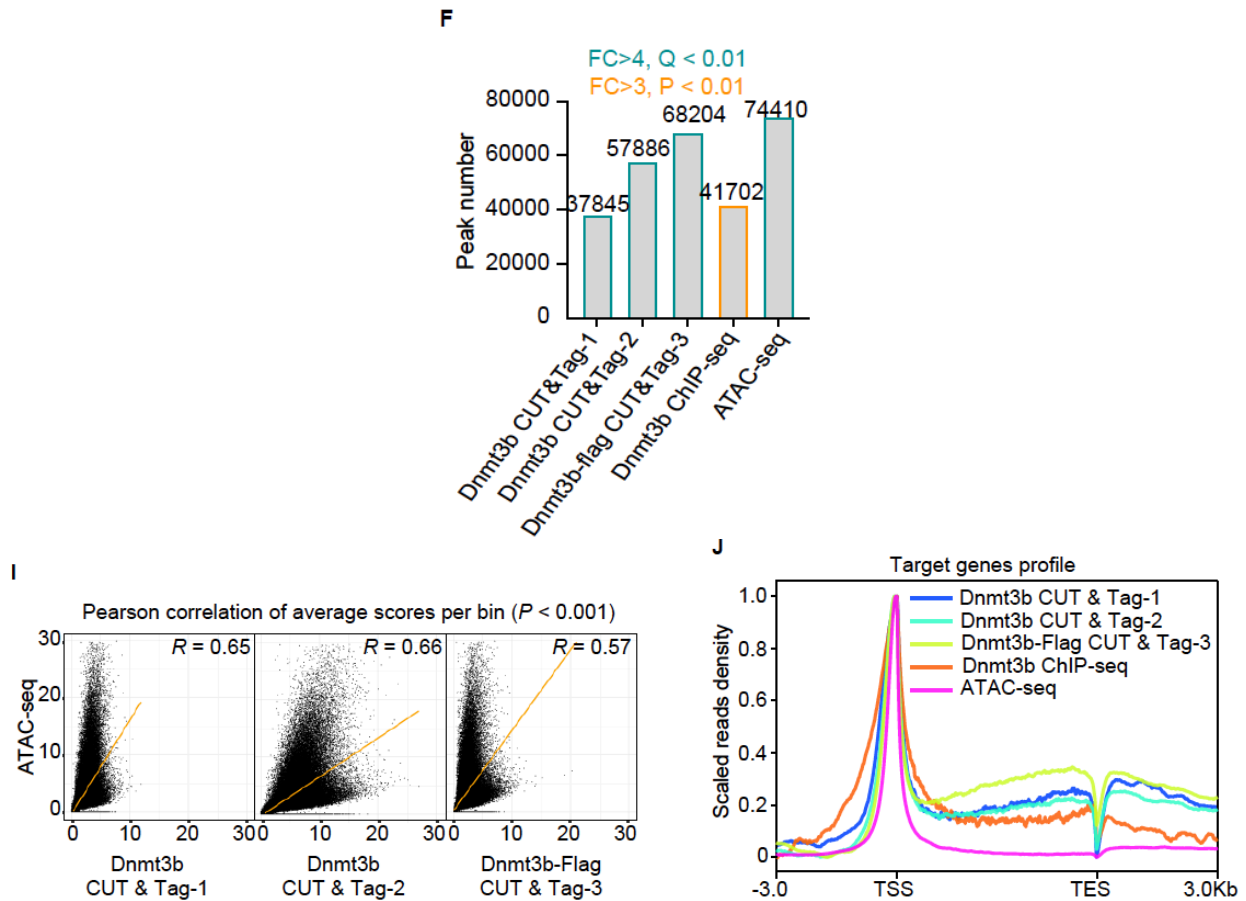
In summary, our four assays for Dnmt3b binding profiles, which utilized two methods (CUT&Tag and ChIP-seq) and two antibodies (anti-Dnmt3b and anti-Flag), are reproducible. However, ChIP-seq exhibited higher background noise compared to CUT&Tag, which provides a higher signal-to-noise ratio, particularly when using spike-in for calibration (CUT&Tag-2) or anti-Flag (CUT&Tag-3). This may explain the observed variability among these assays by visual inspection, as the reviewer pointed out below, particularly why “the Dnmt3b CUT&Tag-2 experiments in Figure 5a do not closely resemble the DNMT3B ChIP-seq data or the Dnmt3b CUT&Tag-1 data”.

I am particularly worried because, by visual inspection, the enrichment intensity for these enzymes over SMAD2/3 peaks appears to be notably different between the epitope-tagged CUT&Tag and the antibody-directed ChIP-seq experiments (see Figures 5a and ED Figure 5a). It has been reported that CUT&Tag can be sensitive to excessive tagmentation under imperfect conditions, leading to artifactual ATAC-seq-like profiles, and this can be especially problematic for factors with broad genomic occupancy such as the DNMT3s (see, for example, PMID: 32913232 for description of these artifacts). To me, all three DNMT3B ChIP/CUT&Tag experiments seem quite variable: the extremely punctate signal within the “Dnmt3b CUT&Tag-2” experiments in Figure 5a don’t resemble the DNMT3B ChIP-seq data or “Dnmt3b CUT&Tag-1” data. All told, these differences raise some caution about the authors’ interpretations and I ask them to reinvestigate which signal is accurate (this could include the generation of corresponding ATAC-seq data to establish that the signals don’t correlate in a manner that would suggest an artifact, as well as further replicate power and/or description/justification of the reaction conditions used).

We thank the reviewer for the valuable suggestion. To exclude potential ATAC-seq-like artifacts from the Dnmt3b CUT&Tag assays, we conducted the following experiments and analyses as recommended by the reviewer.

1) We generated corresponding ATAC-seq data in WT EBs and identified 74,410 peaks using the same criteria as in CUT&Tag (fold change ≥ 4 , $Q \leq 0.01$, Revised Supplemental Figure 5F). The correlations between CUT&Tag and ATAC-seq signals, with R values ranging from 0.57 to 0.65, were lower than the high reproducibility among Dnmt3b CUT&Tag biological replicates, with R values ranging from 0.89 to 0.93 (Revised Supplemental Figure 5I).

2) We analyzed signal distributions across 14,461, 16,684, and 17,170 target genes from three Dnmt3b CUT&Tag assays, 14,751 target genes from the Dnmt3b ChIP-seq assay, and 18,120 target genes from the ATAC-seq assay (Revised Supplemental Figure 5J). Dnmt3b binding profiles exhibited higher similarity, particularly among CUT&Tag assays (Revised Supplemental Figure 5J). In contrast to the ATAC-seq assay, all Dnmt3b binding profiles of CUT&Tag or ChIP-seq showed greater enrichment in gene body regions, consistent with published data². All these findings further validate the accuracy of the Dnmt3b binding patterns.



Supplementary Figure 5. The genome-wide binding profiles of Dnmt3b.

F. Bar plots showing the number of Dnmt3b binding peaks identified by MACS2 in three CUT & Tag assays and one ChIP-seq assay, as well as ATAC-seq peaks in D3 EBs. CUT & Tag peaks and ATAC-seq peaks were called with a fold change cutoff of 4 and a Q-value of 0.01 (Cyan), while ChIP-seq peaks used a fold change cutoff of 3 and a P-value of 0.01 (yellow).

I. Scatterplots showing the Pearson correlation (R) of signal intensity were compared between CUT&Tag assays for Dnmt3b binding profiles and the ATAC-seq assay across the whole genome, which were divided into 1 kb bins. An F-test was performed. R values were shown as indicated.

J. Normalized signal density profiles were across 14,461, 16,684, and 17,170 target genes for three Dnmt3b CUT&Tag assays, 14,751 target genes for the Dnmt3b ChIP-seq assay, and 18,120 target genes for the ATAC-seq assay. TSS, transcription start site; TES, transcription end site.

I'm also surprised that the authors elected to go with a C-terminal Flag tag, as this would be situated proximally to the catalytic domain. Could the authors provide additional data or referenced validation that these epitope-tagged versions are functional, as well as further supporting data that the integrated transgene is correct (currently we only have ED Fig. 5b for reporting these lines)?

We appreciate the reviewer for raising this important point.

1) In a 2023 paper published in *Nature Cell Biology*³, the authors introduced a Degron tag (FKBP12^{F36V}-FLAG-HA-P2A-EGFP) into the C-terminus of the endogenous DNMT3A locus in human embryonic stem cells (hESCs) and a C-terminal single flag-tag into the endogenous Dnmt3a gene in mouse embryonic stem cells (mESCs), as shown in Figure 4a and Extended Data Figure 8 of the paper, respectively³. Using these tagged Dnmt3a cell lines, they studied the function of Dnmt3a in various aspects of hESC and mESC biology³.

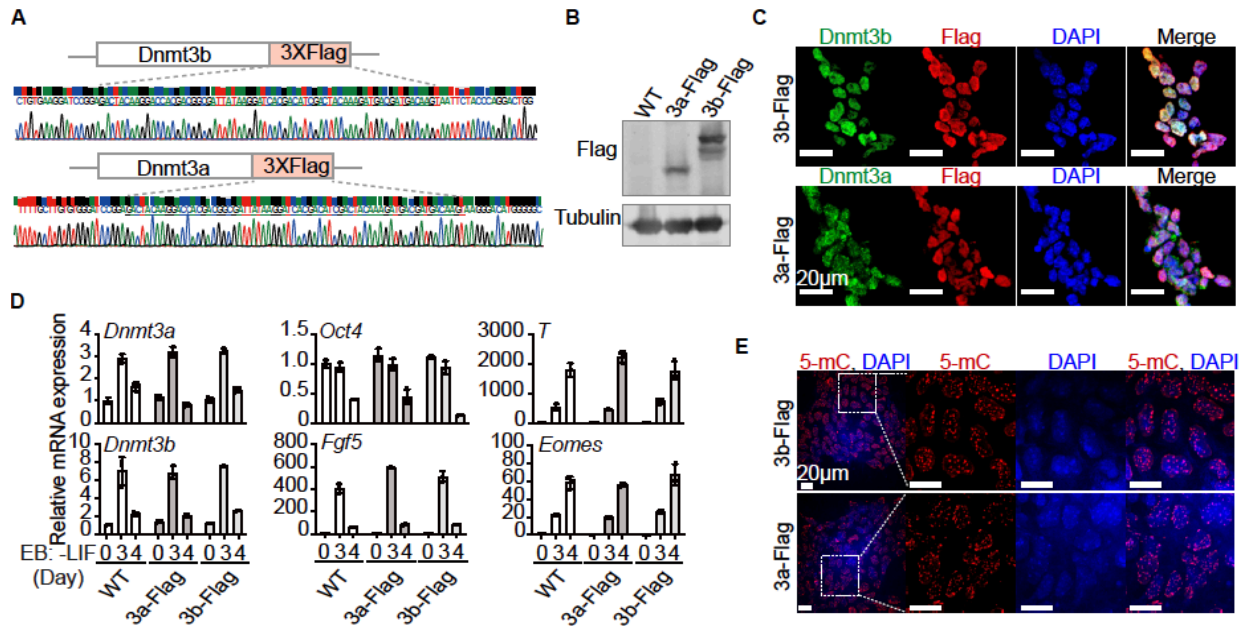
[REDACTED]

(PMID: 37024683) Figure 4a, [REDACTED]

[REDACTED]

(PMID: 37024683) Extended Data Fig. 8 [REDACTED]

2) As the reviewer suggested, we conducted more experiments to provide data that the integrated transgene is correct. These clones were sequenced and validated by both western blotting and immunofluorescence (Revised Supplemental Figure 5A-C). Furthermore, the cells formed EBs correctly and showed upregulation of primed and differentiation genes upon LIF removal, as demonstrated by qRT-PCR (Revised Supplemental Figure 5D). They also exhibited high levels of 5-mC, comparable to WT EBs (Revised Supplemental Figure Figure 5E).



Supplementary Figure 5. The genome-wide binding profiles of Dnmt3b.

- A schematic representation of 3xFlag-tag insertion at the C-terminus of Dnmt3b (3b-Flag) or Dnmt3a (3a-Flag) is shown with the corresponding sequencing data displayed below.
- Western blot analyzed Flag-tagged proteins expression in 3a-Flag and 3b-Flag EBs (bottom). Tubulin serves as a loading control.
- Representative confocal microscope images of Dnmt3a or 3b (green), Flag (red) and DAPI (blue) staining in 3b-Flag and 3a-Flag EBs. Scale bars, 20 μ M.
- The mRNA levels of Dnmt3a, Dnmt3b, Fgf5, T, Eomes and Oct4 during EB differentiation (D0, D3 and D4) in WT, 3a-Flag and 3b-Flag were analyzed by qRT-PCR.
- Representative confocal microscope images of 5-mC (Red) and DAPI (blue) staining in 3b-Flag and 3a-Flag EBs. Scale bars, 20 μ M.

Finally, please refine the model Figure 6 such that DNMT3B does not appear to be interacting with H3K4 methylated (or H3K27ac) chromatin, as this is canonically an inhibitory interaction that blocks DNMT3 binding or de novo activity. I similarly find the conjecture that these interactions are likely to be H3K36me3 related to be a bit overstated (at the very least, the DNMT3 PWWP domains can also interact with H3K36me2) but can understand if the authors would like to keep this interpretation given their rescue experiments with mutant Dnmt3b constructs.

We apologize for the lack of clarity and have refined the model in Figure 6. We propose that Dnmt3b interacts with Smad2/3 to facilitate their binding to the hypomethylated centers of promoters and enhancers of epiblast genes, likely in a manner dependent on H3K36me3. However, we fully agree with the reviewer that we cannot rule out the involvement of other mechanisms including H3K36me2, which requires further study in the future.

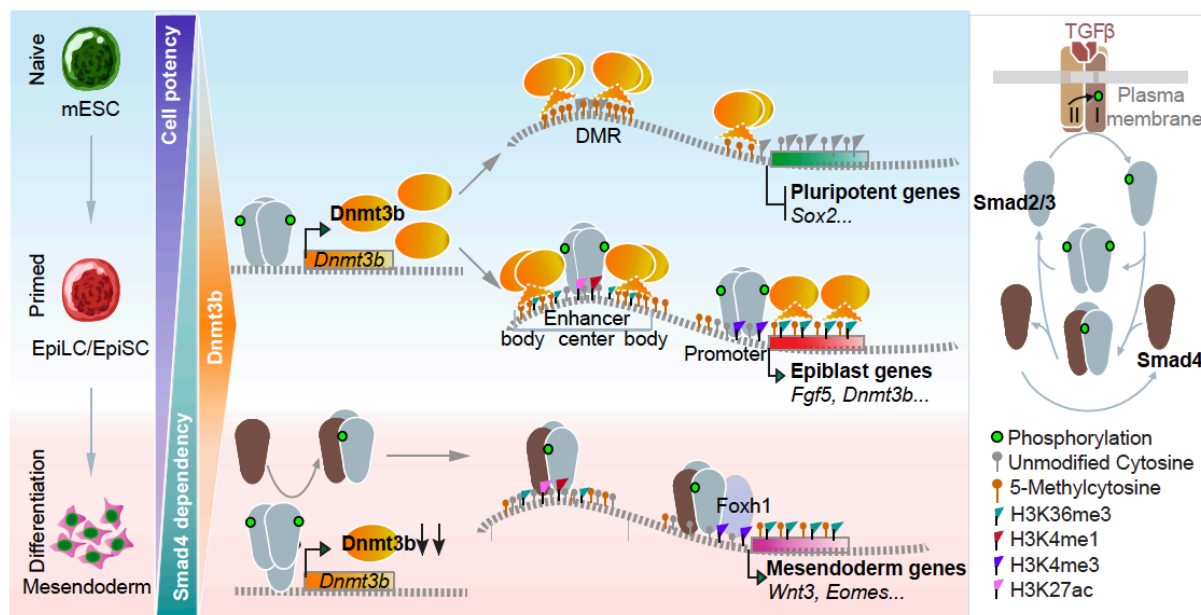


Figure 6. A two-step model describes the transition from naïve to primed epiblast and subsequent differentiation.

Reviewer #2 (Remarks to the Author):

The revised paper NCOMMS-23-43854A by Zhao and colleagues describes a novel mechanism of the early steps of mouse embryonic stem cell development ex vivo that starts with pluripotent stem cells, proceeds towards the epiblast stage during a naïve to primed transition step and ends with mesendodermal differentiation. Two key factors are studied, the signaling factors Smad2, Smad3 and Smad4 of the TGF-beta family pathways (represented by nodal or activin at early embryonic stages) and the epigenetic factor Dnmt3b that methylates cytosines on CpG-rich DNA sequences. Smad2, Smad3 and Smad4 are known to function as a heterotrimer in a variety of physiological and pathological contexts. The paper presents three tiers of novel mechanisms: a) the early stage naïve to primed transition is driven by Smad2 and Smad3 but Smad4 is dispensable for this biological process. During this stage pluripotency factors like Oct4 and Sox2 are downregulated and Dnmt3b levels are enhanced, the latter contributing to DNA methylation and silencing of the pluripotency genes. b) Mesendodermal differentiation follows, which requires the established Smad2/Smad3/Smad4 transcriptional complex. c) Smad2 and Smad3, but not Smad4, are shown to interact with Dnmt3b, and such complexes position the Smads on hypomethylated gene regions, thus promoting gene expression by the cooperative signal of Smads and Dnmt3b, a protein otherwise linked with gene repression. The positive transcriptional action of Smad2/3-Dnmt3b is also shown to rely on the interaction with histone 3, trimethylated at lysine 36. The paper covers a vast array of experimental methods and well-controlled experiment. The revised paper conveys a clear and novel molecular mechanism that is investigated using a plethora of genome-wide and proteomic technologies and an impressive array of genetic interventions followed by rescue experiments with normal and mutant proteins. The revised paper has enhanced significantly the mechanistic understanding of the Smad2/3-Dnmt3b complex and its chromatin-based association with target genes, on which, this complex acts as a positive regulator of transcription. The paper is succinctly written and also provides a large resource of chromatin-related, protein-related and phenotypic data-sets that can stimulate many new studies in basic stem cell and developmental biology. The authors made an impressive effort to respond to the comments of the reviewers and have delivered an outstanding article.

We greatly appreciate the reviewer for the positive feedback and for highlighting our interesting novel findings. The reviewers' insightful suggestions have greatly enhanced our understanding and interpretation of the data in this work.

Based on this latter comment, I recommend the following and final (from my side) minor corrections.

As suggested by the reviewer, we have corrected the following typos and errors.

Typographical and minor errors:

1. Line 51: the Smad2-Smad3 is also a heterotrimer. Please correct the term "homotrimer" that implies Smad2-Smad2-Smad2 or Smad3-Smad3-Smad3.

We have revised it to: Smad2 and Smad3 (Smad2/3) are phosphorylated and form a trimer, either with or without the common mediator Smad4.

2. Line 61: Smad4 with capital S.

We have revised it to: During pancreas development, Smad2/3, but not Smad4, plays a key role.

3. Line 106: Smad instead of SMAD as you use the mouse nomenclature throughout your paper.

We have revised it to: Nodal-based Smad signaling-transduction.

4. Line 117: Although mESCs depleted from (add the word: from).

We have added the word "from".

5. Line 235: ..., we were curious... or ..., we became curious..., but not ..., we are curious...

We have revised it to: Since Smad4 is not essential for the transition of mESCs to EpiLCs, we were curious about its importance in epiblast-derived differentiation.

6. Line 324: ...for epiblast gene (not genes) expression...

We have deleted the letter "s".

7. Line 336: ...remains constant (not consistent) ...

We have revised it to: However, the protein level of Dnmt3a remains constant from EpiLC to mesendoderm stage.

8. Line 536: It can suppress gene transcription, ...: It refers to DNA methylation, but this is not obvious. Better to be explicit: DNA methylation can suppress gene transcription, ...

We have revised it to: DNA methylation can suppress gene transcription, especially in CpG islands of promoters or in enhancers with repressive histone modifications.

9. Line 848: if possible spell the last name correctly: J. Massagué (with an accent on the last e).

We have added the accent to the final 'e' in Massagué.

10. Lines 860-861: performed the bioinformatic analysis. (not the bioinformatics).

We have deleted the letter "s".

11. The references do not use a single and consistent style when it comes to the journal name.
We have updated the references.

Reference

1. Kaya-Okur HS, Janssens DH, Henikoff JG, Ahmad K, Henikoff S. Efficient low-cost chromatin profiling with CUT&Tag. *Nat Protoc* **15**, 3264-3283 (2020).
2. Baubec T, *et al.* Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**, 243-247 (2015).
3. Ramabadran R, *et al.* DNMT3A-coordinated splicing governs the stem state switch towards differentiation in embryonic and haematopoietic stem cells. *Nat Cell Biol* **25**, 528-539 (2023).

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

I thank the authors again for their work in stabilizing specific claims regarding DNMT3B and TGFb in the naïve to primed transition, including more detailed characterization/validation of their flag-tagged mESC lines and deeper analysis of their CUT&Tag and ChIP-seq data. Ultimately I still find many of the broad conclusions regarding Smad2/3 binding, DNMT3B upregulation and the naïve-to-primed epiblast transition to be compelling and suitable for publication. Conversely, I still believe the CUT&Tag/native ChIP-seq data is quite problematic and in contrast with what is broadly understood about DNMT3 binding and function, particularly during the Naïve to Primed transition where DNMT3B is understood to be performing a global role in remethylating the preimplantation genome. To me, the overlapping genomic occupancy between SMAD2/3 and DNMT3B are likely an artifact of imperfect enrichment techniques and biasing towards nucleosome-depleted euchromatin when IP'ing globally distributed factors. To me, this issue is not easily rectified with additional experiments, but I think it would be sufficient to publish if the language of this last section and the discussion are more carefully worded.

We greatly appreciate the reviewer for the positive feedback and insightful suggestions, which have greatly enhanced our understanding and interpretation of the data in this work. I agree that, at this point, we cannot rule out the possibility of technical issues raised by the Dnmt3b CUT&Tag analysis, particularly regarding its binding to the proximal promoter region (PP). But these data may also capture true signals, such as well-known gene-body enrichment. As the reviewer suggested, we have carefully revised the wording in the final section and added an extra paragraph in the discussion to address these concerns. All these points suggest that improved technologies or methods may be required for future studies. Please find our point-by-point response to the reviewer's comments below.

The authors show in their newly provided figures that their tagmentation-based CUT&Tag data does not overlap with their sonication-based ChIP-seq data (ED Figure 5g) and that this data shows a strong linear relationship with normal ATAC-seq data (ED Figure 5I). Yes, within these scatterplots, the average intensity is higher for the ATAC-seq than it is for DNMT3B CUT&Tag for the majority of peaks, but this is the precise definition of an artifact (a signal that can be

explained by a technical detail of the assay, namely the use of Tn5 transposase). Similarly, it is not appropriate for the authors to comment on this assay's improved signal to noise ratio around peaks for regulators that are not expected to generally have peaks, but are rather broadly distributed over the genome (like the DNMT3's).

To me this concern is particularly born out by the correspondingly high normalized signal intensity around the TSS in ED Figure 5J, an enrichment that is very hard to rectify against DNMT3's allosteric inhibition and poor correlation with promoter-enriched H3K4me3 (see previous reports such as Baubec et al, 2015, particularly Figure 3 and ED Figure 2 of that paper that show ¹). Yes, ED Figure 5J shows that there is gene-body enrichment in the CUT&Tag vs ATAC-seq, which appears to be true signal. In contrast I would presume the authors agree that the TSS enrichment is an "artifact" compatible with basal tagmentation of euchromatin. Given that the SMAD binding sites are likely to be similarly sensitive to Tn5, how are we supposed to distinguish this non-canonical and somewhat exotic model of cooperative DNMT3B binding from an artifact?

I do not think it's fair to ask the authors to rectify this challenging aspect of their manuscript with more experiments, particularly after several rounds of review where the explicit biochemical mechanism has been a sticking point. Instead, I ask that they soften the language in the last section of the manuscript, including language that insists that the CUT&Tag data is better based upon signal to noise ratio, and to comment on which features of ED Figure 5 the authors believe to be true or not based upon the literature.

We appreciate the reviewer for raising this important point.

1) We have carefully revised the wording in the final section and removed all assertions that the CUT&Tag data is superior based on the signal-to-noise ratio. Additionally, we have reorganized the description comparing ATAC-seq, CUT&Tag, and ChIP-seq, and then directly addressed the concerns as follows:

"Although Dnmt3b binding to enhancer and GB is consistent with previous reports^{1, 2}, its binding to PP was unexpected (Figure 5A and Supplementary Figure 5J), since earlier studies confirmed the absence of Dnmt3b around transcription start sites (TSS)¹. At present, we cannot rule out the

possibility that this could be potentially caused by biased binding towards nucleosome-depleted euchromatin generated by the CUT&Tag³. “

2) We have further elaborated on this concern in the discussion section, as shown in the subsequent section.

Then, I would ask them to add a paragraph to the discussion about how it remains unclear precisely how DNMT3B may be molecularly interacting with the Nodal signaling pathway during mESC differentiation, including an acknowledgement that it is difficult to distinguish whether or not the broader ChIP-seq data (and findings from papers such as Baubec et al.) and more punctate CUT&Tag data disagree because of the improved nature of CUT&Tag or because it is not an ideal assay to study regulators with global functions. Please also highlight in this paragraph the challenging kinetics and cellular heterogeneity inherent to stem cell-based differentiation assays, especially given the rapid and possibly overlapping transitions between naïve to primed and primed to mesendoderm. The conclusions that are most appropriate are that SMAD2/3 induces DNMT3B, DNMT3B & SMADs appear to biochemically interact during a window where they are both active and highly expressed, and this window precedes co-binding by SMAD4. If the authors appropriately highlight the current caveats as room for future investigation, they are welcome to maintain Figure 6 as a “potential” or “proposed” model.

We deeply appreciate the reviewer's guidance in improving our manuscript. As suggested, we have added the following paragraph discussing the remaining challenges and caveats in this work that require further study.

“It is worth noting that Dnmt3b exhibited a broad binding pattern in ChIP-seq data but displayed a more punctate pattern in CUT&Tag data. This discrepancy makes it difficult to ascertain whether the difference arises from the improved resolution of the CUT&Tag method or if it indicates that CUT&Tag may not be the most suitable assay for studying regulators, especially those with global functions. While Dnmt3b binding to enhancers and GB aligns with previous reports^{1,2}, its binding to PP regions contradicts its known absence from TSS of actively transcribed genes and is inconsistent with its poor correlation with promoter-enriched H3K4me3¹. Currently, we cannot rule out the possibility that these observations are influenced by Tn5-directed bias toward binding in open chromatin regions during the CUT&Tag assay. Future studies may require improved technologies or methods. Given the challenging kinetics and cellular heterogeneity inherent to

stem cell-based differentiation assays, as well as the rapid and potentially overlapping transitions from naïve to primed pluripotency and subsequent differentiation, clarifying how Dnmt3b precisely interacts with and facilitates the Nodal signaling pathway remains difficult and requires further investigation.”

Please also more generally refer to H3K36 methylation (or me2/3) throughout the manuscript rather than H3K36me3, which is generally associated with transcriptional elongation and not enriched at enhancers.

We appreciate the valuable suggestion and have revised "H3K36me3" to "H3K36me2/3" in the manuscript.

Reference:

1. Baubec T, *et al.* Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**, 243-247 (2015).
2. Rinaldi L, *et al.* Dnmt3a and Dnmt3b Associate with Enhancers to Regulate Human Epidermal Stem Cell Homeostasis. *Cell stem cell* **19**, 491-501 (2016).
3. Kaya-Okur HS, Janssens DH, Henikoff JG, Ahmad K, Henikoff S. Efficient low-cost chromatin profiling with CUT&Tag. *Nat Protoc* **15**, 3264-3283 (2020).