THE
EMBO
JOURNAL

Manuscript EMBO-2012-83939

Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm

Irene Aksoy, Ralf Jauch, Jiaxuan Chen, Mateusz Dyla, Ushashree Divakar, Gireesh K. Bogu, Roy Teo, Calista Keow Leng Ng, Wishva Herath, Sun Lili, Andrew P. Hutchins, Paul Robson, Prasanna R Kolatkar and Lawrence W. Stanton

Corresponding author: Lawrence W Stanton, Genome Institute of Singapore

Review timeline:

Submission date:	16 November 2012
Editorial Decision:	13 December 2012
Revision received:	23 January 2013
Accepted:	24 January 2013

Transaction Report:

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

Editor: Thomas Schwarz-Romond

1st Editorial Decision 13 December 2012

Thank you very much for submitting your study on Oct4-partnering with Sox2/Sox17 to govern alternative developmental programs for consideration to The EMBO Journal editorial office.

The attached comments reflect on the scientific merits of your study. The referees also offer constructive support to further substantiate the very interesting proposal and improve the quality/comprehension of the paper.

Specifically, it would be interesting to experimentally address ref#1's first point and assess the consequences of a single base insertion into the compressed cis-element.
Furthermore, ref#2 brings up a number of very valid points that need your attention, though I would not necessarily insist on expanding the study further to Sox7 (ref#2, point3); You are of course welcome to add these, in case such data might already be at hand OR relatively easily to be acquired.

Based on these, I am delighted to invite submission of a thoroughly revised study for final assessment by some of the original referees.

Please do not hesitate to get in touch in case that further questions arise (preferably via E-mail).

Finally, I do have to formerly remind you that The EMBO Journal considers only one round of revisions with the ultimate decision on publication being solely dependent on the outline and

strength of the ultimate manuscript.

REFEREE REPORTS:

Referee #1:

In this work, the Stanton lab expand on their previous findings relating the binding of Oct4 with the distinct Sox family members, Sox2 and Sox17 (Jauch et al, 2011, Stem Cells) to examine transcription factor binding in the context of chromatin and the effects of the underlying compressed or canonical Oct/Sox motifs on cell-specific function of enhancers. This is an important study that merits prominent publication as it instructs us on how enhancer specificity for ES cells or primitive endoderm can be modified by small characterised alterations in the interaction surfaces of transcription factors that co-operate in DNA binding in order to interpret the cell specific instructions embedded in the enhancer sequences.

I do have a few comments that the authors may wish to consider that could enhance their manuscript. Several of these relate to figure legends which are not always explicit or intelligible.

1 Can Sox-regulated gene expression be switched from endoderm to ES cell specific by base insertion into compressed motifs to make canonical motifs? This could be done by modifying the experiments in Fig 8C. These include some "mutants" of the Oct/Sox motif but these were not explicitly described in the paper and they should be. However, the more exciting aspect of this would have been the inclusion of a single base insertion into the compressed motifs of these regulatory elements to test the clear prediction of the author's data that this could switch interpretation of the enhancer code and induce undifferentiated cell specificity.

2 I found the discussion of the data in paragraph 2 of page 6 rather circular. It is obviously no surprise that the authors found that "Sox2 is more likely to co-bind with Oct4Sox2 than with Oct4Sox17", since Oct4Sox2 sites are those in which the authors found Sox2 bound. I think the authors should find a clearer way of describing their findings.

3 Fig 2A is deeply confusing. What are yellow and purple circles and just exactly how do they relate to the bar chart below? Authors should make it clearer what the pink and black bars are.

4 Fig 5E details should be provided of how this was achieved as this IP is not trivial to show and has not been demonstrated in many other studies on Oct4 and Sox2.

5 Page 11, line 11 reference to Sox motif being the top scoring rather underplays the absence of the compressed motif. This seems important. Given the authors previous data (Jauch et al., 2011) showing 17EK cannot bind the compressed motif but that S2KE does not bind the compressed motif as effectively as S17 in their previous EMSAs, this is worthy of more explicit comment.

6 Immediately after this the authors should comment on whether the >60% canonical for O/S17EK compared to >50% for O/S2 means that S17EK is better at binding the canonical motif than Sox2 and whether it drives pluripotent gene expression better.

7 page 13. I cannot see how the authors can conclude that "Oct4 is required for differentiation towards an extra-endodermal cell fate and that the physical interaction of Oct4 and Sox17 in F9 cells is also required" from the experiments presented. Oct4 may be required to allow the cells to survive to the point where endoderm differentiation becomes possible; is there an Oct4 knock-out in F9 cells that backs up the authors statement? The authors should consider re-wording more restrained manner.

Minor points

8 Fig1A This is unclear and should be re-drawn in a more informative cartoon style; where is rtTA expressed from?

9 Fig1B confusing and should be presented side by side, Sox2 and Sox17 with the relevant actin

probing below.

10 Fig1C axis is relative to what?

11 Fig1D what antibodies are in use here? Is this a reprobing? Why is there only 1 actin control?

12 Fig 3C what are the black horizontal bars between the sky blue element and the motif density plot? What are the numbers under the motif density indicating?

13 Where is the Fig 3G referred to in the legend?

14 Figure 4B and C "no motif" should be replaced by "no composite motif" to maintain consistency with Fig 2 and to underline the fact that a Sox motif is found (Fig 4A(ii)).

15 Fig 6 A can this be altered to make the numbers legible?

16 Fig 7B this is too small to judge. The orientation of the blastocysts should be consistent between Oct4 and Sox17 panels.

17 Fig 7 C what does "ratio E3.75/ICM" mean?

18 Figure 8 the reference to Kawakami is missing in the reference list.

19 Page 8 find another way of referring to "shadow" motifs to avoid confusion with "shadow enhancers" (Hong et al. Science, 2008, 321, 1314).

20 On page 11 reference to Fig 2B(i) on line 8 should be to 2B(ii).

Referee #2:

The manuscript by Askoy and colleagues describes Oct4 switching genomic occupancy with pluripotent cell differentiation into extraembryonic endoderm suggesting Oct4 drives alternative developmental programs. Interestingly this correlated with Oct4 changing genomic occupancy from an Oct4/Sox2 DNA binding motif (canonical motif) to an Oct4/Sox17 motif (compressed motif). Overall the manuscript is interesting. I do however have some points that need to be addressed before recommending publication.

Major points:

1. The paper can be improved by combining the microarray analysis of Sox17 OE cells with ChIP-Seq data obtained in the first part of the manuscript. Particularly,

a) How many and which of the endoderm-specific genes that have a compressed Oct/Sox motif are upregulated in Sox17 OE cells and how many and which were demonstrated by ChIP-Seq to have both Oct4 and Sox17 peaks on them?

b) Canonical Oct/Sox motif was detected in many genes downregulated upon Sox17 OE. Is Oct4/Sox2 binding reduced or lost on these motifs upon Sox17 overexpression? This would confirm authors' hypothesis that Oct4/Sox2 pair act as a transcription activator of pluripotency genes and once Sox17 is introduced Oct4 is relocated to endoderm-specific genes by Sox17 and pluripotency genes get repressed while endoderm genes get activated.

2. In the RA differentiation protocol more markers of different lineages should be analysed to confirm that these conditions favour endoderm formation. Also, established XEN lines or other endoderm sample should be used as a positive control for the gene expression analysis in Figure 5B to appreciate the acquisition of the endoderm cell fate.

3. Another member of Sox family, Sox7, is a known marker of primitive endoderm. Can Oct4 also cooperatively bind with this transcription factor? If yes, what motif is used and whether it is enriched in endoderm-specific genes?

4. It is not clear from the text when Oct4 was knockdown in Figure 5F? Throughout the paper authors state that Oct4 is essential for differentiation towards extraembryonic endoderm. However, if this is based on Oct4 knocdown in ES cells then this is known to result in differentiation into trophectoderm. This means it has nothing to do with a specific block in differentiation into extraembryonic endoderm. Oct4 expression is required for the maintenance of pluripotency so the knockdown of Oct4 should be performed at a stage just prior Sox17 activation and Sox2 donwregulation which according to the authors is day 2 after RA treatment.

5. Oct4 is not expressed in XEN cells or primitive endoderm of the embryo. Is Sox17 binding pattern is different in XEN cells than in EC cells at early stages of endoderm differentiation?

6. Figure 6c. The authors state that Sox17 is necessary to recruit Oct4 in RA differentiation. To be able to state this authors need to demonstrate by SiRNA against Sox17 that Oct4 RA peak is lost.

7. In Figure 8c authors should also provide reporter expression of enhancers containing canonical motif as a negative control. Flow cyitometry quantification before and after differentiation is also required.

Minor points:

1. Is the data in Figure 1C normalized to the untreated control? What is the fold degree of overexpression of Sox2 and Sox17 in the generated lines?
2. Figure legend 2K is missing and the following ones are shifted up.

Referee #3:

The authors provide an elegant and important manuscript elucidating how alternative partnering of Oct4 with Sox2 or Sox17 regulates endodermal specification from pluripotent cells. The authors identify and prove the functional importance of a new "compressed" enhancer motif that is targeted by Oct4/Sox17 complexes. The authors also provide conclusive Sox17 mutation swap analysis with Sox2 that is impressive. They also used an embryonal carcinoma model to identify, based on the compressed motif, novel genes upregulated upon endodermal differentiation, and show that their upregulation is dependent both on the presence of Oct4 and Sox17 during this process.

The manuscript is well written with appropriate references. The conclusions are well supported by the data. I have the following minor points:

1) Throughout the manuscript, the authors use pompous language regarding how they prove the "enhancer code". The manuscript would benefit from toning down the novelty and focus on clarifying the conclusions.

2) It would be elegant to include Sox2 staining in vivo in early and late ICMs, relative to sox17 (shown in Figure 7B).

3) The authors should consider revising or omitting their single cells reanalysis presented in Figure 7a. It appears that both extremes of phenotype (Epi-like and Pre-like), have similar Nanog, Sox2 and Gata4 expression levels?. I do not understand how they can classify the cells into two groups distinctly based on the expression patterns and curves shown.

© EMBO

1st Revision - authors' response 23 January 2013

We have addressed the comments from all 3 reviewers and provide our detailed responses to each below. We are grateful that the reviewers were generally positive about our findings and the quality of our data. They have also provided thoughtful suggestions on how we can improve our paper. For the major points raised by the reviewers, we have provided additional experimental results, added new data, or revised our presentation in a manner that we believe addresses each concern. Minor points were easily rectified throughout.

Detailed Responses to Reviewer comments:

Referee #1:

In this work, the Stanton lab expand on their previous findings relating the binding of Oct4 with the distinct Sox family members, Sox2 and Sox17 (Jauch et al, 2011, Stem Cells) to examine transcription factor binding in the context of chromatin and the effects of the underlying compressed or canonical Oct/Sox motifs on cell-specific function of enhancers. This is an important study that merits prominent publication as it instructs us on how enhancer specificity for ES cells or primitive endoderm can be modified by small characterised alterations in the interaction surfaces of transcription factors that co-operate in DNA binding in order to interpret the cell specific instructions embedded in the enhancer sequences.

I do have a few comments that the authors may wish to consider that could enhance their manuscript. Several of these relate to figure legends which are not always explicit or intelligible.

1 Can Sox-regulated gene expression be switched from endoderm to ES cell specific by base insertion into compressed motifs to make canonical motifs? This could be done by modifying the experiments in Fig 8C. These include some "mutants" of the Oct/Sox motif but these were not explicitly described in the paper and they should be. However, the more exciting aspect of this would have been the inclusion of a single base insertion into the compressed motifs of these regulatory elements to test the clear prediction of the author's data that this could switch interpretation of the enhancer code and induce undifferentiated cell specificity.

We now describe in more detail the "Mutants" used for Figure 8C in Supplementary Table S5 and added, as suggested by Referee #1, new "Canonical mutants" which are compressed motifs in which we inserted a single nucleotide to transform them into canonical motifs. We tested them in undifferentiated and differentiated F9 cells. Results are presented in Figure 8 and Supplementary Figure 6. We now modify the text p19: "Moreover, when we introduced rational mutations that degenerated the compressed motif of the described enhancers by the replacement of multiple critical nucleotides within the Sox and Oct half-sites (degenerating mutants) reporter activity upon differentiation was completely abolished (Figure 8C(ii)). Next, to more rigorously assess the relevance of the compressed motif for reporter expression we generated mutations by inserting a single nucleotide between Sox and Oct half-sites to transform the compressed motif into a "canonical motif" (canonical mutants, Figure 8C(iii), Figure S6). Notably, even this subtle disturbance to architecture of the enhancer led to a complete loss of the expression of GFP in differentiating F9 cells. This indicates that cooperative binding of Sox17/Oct4 to the compressed motif is indeed necessary to drive the expression of these endodermal genes. In a reverse experiment we also tested the "canonical mutants" in undifferentiated F9 cells and observed a gain in GFP expression for Sall4 and Smad2 driven reporters that would otherwise have been silent (Figure 8C (iv), Figure S6). These results indicate that a single nucleotide insertion can dramatically alter enhancer activities by recruiting alternative Sox/Oct4 complexes."

2 I found the discussion of the data in paragraph 2 of page 6 rather circular. It is obviously no surprise that the authors found that "Sox2 is more likely to co-bind with Oct4Sox2 than with

Oct4Sox17", since Oct4Sox2 sites are those in which the authors found Sox2 bound. I think the authors should find a clearer way of describing their findings.

To clarify, 'Oct4^{Sox2}' refers to Oct4 ChIP-seq in the context of the exogenous expression of Sox2, and Oct4^{Sox17} refers to Oct4 ChIP-seq in the context of exogenous Sox17 expression. Hence it wasn't that obvious that Sox2 is more likely to preferentially intersect with 'Oct4^{Sox2}' as compared with 'Oct4^{Sox17}'. We now rephrase the main text to avoid confusion p6: "We first identified Oct4 sites (i) in the context of exogenous Sox2 expression (Oct4^{Sox2}) and (ii) Oct4 sites in the context of exogenous Sox17 expression (Oct4^{Sox17}). Next we intersected these two Oct4 datasets with Sox2 and Sox17 datasets. We found that Oct4^{Sox2} (Oct4 ChIP-seq in Sox2 expressing cells) is more likely to co-bind genomic loci with Sox2 when compared with Sox17. Conversely, Oct4^{Sox17}, (Oct4 ChIP-seq inSox17 expressing cells) is more likely to co-bind with Sox17 when compared to Sox2 (Figure 2A). This indicates that a substantial fraction of Oct4 binding sites is dependent on whether Sox2 or Sox17 is expressed. Hence, changes in the Sox partner lead to a partial genomic re-distribution of Oct4. That is, Sox2/Oct4 co-select a distinct set of genomic loci which is different from the set co-selected by Sox17/Oct4."

3 Fig 2A is deeply confusing. What are yellow and purple circles and just exactly how do they relate to the bar chart below? Authors should make it clearer what the pink and black bars are.

We modified Figure 2A in order to make it clearer.

4 Fig 5E details should be provided of how this was achieved as this IP is not trivial to show and has not been demonstrated in many other studies on Oct4 and Sox2.

Details were added on the Oct4/Sox2 and Oct4/Sox17 co-IP protocol in the Material and Methods section: "F9 cells treated with RA were first incubated with 2 mM of the cross-linking agent DSP (Pierce Biotechnology) for 2 hr at 4°C and then lysed in RIPA buffer (50mM Tris pH 7.5, 150mM NaCl, 5mM EDTA and 0,2-0,5% NP40) supplemented with EDTA-free pellet protease inhibitor (Roche Diagnostics) and 1mM of PMSF. The lysates were then spun and supernatants were used for immunoprecipitation. Lysates were incubated with 5µg of Oct4 antibody (Santa Cruz Biotechnology SC-8628) overnight at 4ºC and then mixed with 25µl of protein G beads (Pierce Crosslink co-IP kit #88805). Protein/Antibody/Bead complexes were finally washed three times with the lysis buffer and kept in protein loading buffer for western-blot. Sox2 and Sox17 western blots were performed using the anti-Sox2 from Santa Cruz Biotechnology (SC17201) and the anti-Sox17 from R&D Systems (AF1924)."

5 Page 11, line 11 reference to Sox motif being the top scoring rather underplays the absence of the compressed motif. This seems important. Given the authors previous data (Jauch et al., 2011) showing 17EK cannot bind the compressed motif but that S2KE does not bind the compressed motif as effectively as S17 in their previous EMSAs, this is worthy of more explicit comment.

We agree and elaborate in the revised text p11: "This is consistent with our previous quantitative EMSA experiments indicating that Sox2KE does not bind on the compressed motif as effectively as Sox17 (Ng et al., 2012). Apparently, Sox2 requires further mutations to engineer Sox17 like DNA recognition in vitro and in ES cells. While simple homology models do not reveal candidate amino acids that need to be modified to install Sox17-like binding activities into Sox2, future structural or molecular dynamics studies of Sox/Oct4 complexes on the compressed motif can potentially resolve this question."

© EMBO

6 Immediately after this the authors should comment on whether the >60% canonical for O/S17EK compared to >50% for O/S2 means that S17EK is better at binding the canonical motif than Sox2 and whether it drives pluripotent gene expression better.

We agree that this important observation deserves further discussion. We now write p12: "Notably, Sox17EK/Oct4^{Sox17EK} sites are more likely to contain canonical motifs than Sox2/Oct4^{Sox2} sites (Fig 2C). This is consistent with our previous finding that Sox17EK cooperates even more strongly with Oct4 than wild-type Sox2 on the canonical motif (Ng et al., 2012). Likewise, Sox17EK has consistently produced a larger number of pluripotent colonies in iPS experiments than Sox2 (Jauch et al, 2011). This suggests that Sox17EK is a more potent dimerization partner for Oct4 on ESC enhancers which leads to an enhanced ability to trigger reprogramming to pluripotency."

7 page 13. I cannot see how the authors can conclude that "Oct4 is required for differentiation towards an extra-endodermal cell fate and that the physical interaction of Oct4 and Sox17 in F9 cells is also required" from the experiments presented. Oct4 may be required to allow the cells to survive to the point where endoderm differentiation becomes possible; is there an Oct4 knock-out in F9 cells that backs up the authors statement? The authors should consider re-wording more restrained manner.

Referee #1 and #2 raised the same concerns. We agree that our conclusion needs to be restrained so we modified the text and now conclude p15: "These results collectively indicate that knock-down of Oct4 impairs the differentiation of F9 cells towards an extra-embryonic endoderm cell fate. These data are supported by a combination of several indirect evidences (i) in zebra fish the Oct4 homolog (spg (pou2) is essential for endoderm (Reim et al, 2004), (ii) Oct4 is robustly expressed in the developing mouse PrE (Guo et al, 2010; Kurimoto et al, 2006; Palmieri et al, 1994), (iii) in ESCs, overexpression of Oct4 results in differentiation towards PrE (Niwa et al, 2000), and finally (iv) XEN-P cells (extra embryonic precursor cells) expressing high levels of Oct4 have been derived from rat blastocysts (Debeb et al, 2009). Altogether these data, with the Oct4 knock-down experiment in F9 cells, suggest that Oct4 is indeed important for the induction PrE differentiation."

Minor points

8 Fig1A This is unclear and should be re-drawn in a more informative cartoon style; where is rtTA expressed from?

We modified Figure 1A and its legend to make it more informative as suggested. As now specified in the legend, rtTA is expressed from the endogenous Rosa26 promoter.

9 Fig1B confusing and should be presented side by side, Sox2 and Sox17 with the relevant actin probing below.

We modified Figure 1B as suggested.

10 Fig1C axis is relative to what?

We modified Figure 1C, expression levels are relative to the corresponding non-treated controls.

11 Fig1D what antibodies are in use here? Is this a reprobing? Why is there only 1 actin control?

We have clarified the legend of Figure 1D. "Proteins were (i) first incubated with the V5 and the Oct4 antibodies (upper panel), (ii) then the membrane was reprobed with the V5 and Sox2 antibodies (middle panel) and (iii) finally with the ß-actin antibody (lower panel)."

© EMBO

12 Fig 3C what are the black horizontal bars between the sky blue element and the motif density plot? What are the numbers under the motif density indicating?

The horizontal bars indicate the presence of canonical/compressed motifs within 50kb of the transcription start sites – the numbers indicate the density using a sliding window. We now clarify the figure legend: (C, D) Microarray probes from KH2 cells expressing Sox17-V5 were ranked from most up-regulated (red) to down-regulated (blue) and the presence of a canonical or compressed motif was scored if found within 50 kb of the TSS of the gene (black horizontal bars indicate the presence of the motif). The density of compressed motif and canonical motif was measured using a sliding window comprising 10% of the total number of microarray probes. Presence of a motif gives the gene a value of 1 and absence of the motif the value 0. A sliding window is then used to generate a density plot. Black dotted line is the mean, and the two read dotted lines are one standard deviation away from the mean.

13 Where is the Fig 3G referred to in the legend?

We mistakenly referred to a Figure 3G, the legend should actually correspond to Figure 2K.

14 Figure 4B and C "no motif" should be replaced by "no composite motif" to maintain consistency with Fig 2 and to underline the fact that a Sox motif is found (Fig 4A(ii)).

We modified Figure 4B as suggested.

15 Fig 6 A can this be altered to make the numbers legible?

We modified Figure 6A by changing the colours and increasing the font size to make the numbers more legible.

16 Fig 7B this is too small to judge. The orientation of the blastocysts should be consistent between Oct4 and Sox17 panels.

We increased the size of the figures and blastocysts are now in the same orientation.

17 Fig 7 C what does "ratio E3.75/ICM" mean?

We modified Figure 7C and replaced "ratio E3.75/ICM" by "E3.75/ICM expression level ratios". We modified the legend accordingly: "Histogram plots representing E3.75 blastocyst versus ICM expression level ratios for genes harbouring a compressed or a canonical motif in their enhancer region. Expression levels were determined by RNA-seq in RPKM (reads per kilo base per million mapped reads)."

18 Figure 8 the reference to Kawakami is missing in the reference list.

The reference to Kawakami has been added.

19 Page 8 find another way of referring to "shadow" motifs to avoid confusion with "shadow enhancers" (Hong et al. Science, 2008, 321, 1314).

We re-phrased the text to: "...suggesting a portion of these motifs comprised motifs found in the proximity of the actual binding site but are not directly bound."

20 On page 11 reference to Fig 2B(i) on line 8 should be to 2B(ii).

The modification has been made in the text.

Referee #2:

The manuscript by Aksoy and colleagues describes Oct4 switching genomic occupancy with pluripotent cell differentiation into extra embryonic endoderm suggesting Oct4 drives alternative developmental programs. Interestingly this correlated with Oct4 changing genomic occupancy from an Oct4/Sox2 DNA binding motif (canonical motif) to an Oct4/Sox17 motif (compressed motif). Overall the manuscript is interesting. I do however have some points that need to be addressed before recommending publication.

Major points:
1. The paper can be improved by combining the microarray analysis of Sox17 OE cells with ChIP-Seq data obtained in the first part of the manuscript. Particularly,

a) How many and which of the endoderm-specific genes that have a compressed Oct/Sox motif are up regulated in Sox17 OE cells and how many and which were demonstrated by ChIP-Seq to have both Oct4 and Sox17 peaks on them?

We agree that this analysis is of crucial importance to support our conclusions. Figure 3C comprises the integration of the ChIP-seq and microarray analysis. This analysis shows that genes unregulated after Sox17 overexpression have a high density of Sox17/Oct4 bound compressed motifs near their TSS, We now discuss this integrated analysis in more depth in the main text p10: "Amongst the up regulated genes containing a Sox17/Oct4 bound compressed motif are genes annotated to play a role during extra embryonic endoderm development such as Col4a1, Col4a2, Lama1, Sall4, Pdgfra, Emb and Hesx1 but also novel candidate regulators such as Tyro3 and Nr2f6."

b) Canonical Oct/Sox motif was detected in many genes down regulated upon Sox17 OE. Is Oct4/Sox2 binding reduced or lost on these motifs upon Sox17 overexpression? This would confirm authors' hypothesis that Oct4/Sox2 pair act as a transcription activator of pluripotency genes and once Sox17 is introduced Oct4 is relocated to endoderm-specific genes by Sox17 and pluripotency genes get repressed while endoderm genes get activated.

We performed the analysis suggested by Referee #1 and found that only 152 of the 2269 Sox2/Oct4 bound canonical motifs are still bound by Sox17/Oct4. Only 4 genes bound by Oct4/Sox17 on a canonical motif were found to be down-regulated after Sox17 OE (Cacng8, Samd4, Sftpd and Rps19). Therefore we decided not to include these data in the manuscript, as it's a very small number and that none of these genes have been shown to be important for pluripotency. Our hypothesis is that Oct4 targets new sets of genes in the presence of Sox17, and that these genes are up-regulated during PrE differentiation of ES or F9 cells which we demonstrate in the manuscript.

2. In the RA differentiation protocol more markers of different lineages should be analysed to confirm that these conditions favour endoderm formation. Also, established XEN lines or other endoderm sample should be used as a positive control for the gene expression analysis in Figure 5B to appreciate the acquisition of the endoderm cell fate.

We added XEN cells in Figure 5B as a control for endoderm induction and found an elevated expression for all endodermal markers. We also analysed the expression of ectodermal (Nestin, Sox1, GFAP) and mesodermal (T, Mixl1, Flk1) markers in F9 cells treated with RA and in XEN

cells, the results are now presented in Supplementary Figure 2.

3. Another member of Sox family, Sox7, is a known marker of primitive endoderm. Can Oct4 also cooperatively bind with this transcription factor? If yes, what motif is used and whether it is enriched in endoderm-specific genes?

We have unfortunately not explored the role of Sox7 in the formation of PrE using a genome-wide ChIP-seq approach, although we agree with Referee #2 that it is potentially interesting. However, in order to help answer this question, we have now provided results of EMSA experiments, presented in Figure 8B, showing the cooperative binding of Sox7 and Oct4 on the compressed motif of the 8 genes selected for the reporter assay in Figure 8. We now modify the text accordingly p19: "We also showed that Sox7 cooperatively binds with Oct4 on the compressed motif (Figure 8B), like Sox17, which can explain the redundancy between Sox7 and Sox17 in regulating endodermal genes".

4. It is not clear from the text when Oct4 was knockdown in Figure 5F? Throughout the paper authors state that Oct4 is essential for differentiation towards extra embryonic endoderm. However, if this is based on Oct4 knockdown in ES cells then this is known to result in differentiation into trophectoderm. This means it has nothing to do with a specific block in differentiation into extra embryonic endoderm. Oct4 expression is required for the maintenance of pluripotency so the knockdown of Oct4 should be performed at a stage just prior Sox17 activation and Sox2 down regulation which according to the authors is day 2 after RA treatment.

This statement that Oct4 is essential for differentiation towards extra-embryonic endoderm is based on the results obtained from F9 cells and not ES cells as the knock-down experiment has been performed in F9 cells. The knock-down was done at d0 and differentiation induced at d1. We chose this experimental setup because the siRNA is really effective only from day2-day3 of transfection which is day1-day2 after differentiation and, as mentioned by Referee #2, Sox17 is activated and Sox2 is down regulated at day2 after RA treatment. For clarity, we have modified Figure 5F and its legend, and added the time points day1 and day2 to Nanog and Oct4 Q-RT-PCR results. We also modified the text corresponding to these data to clarify the experimental setup p14: "We performed a knock-down of Oct4 or Nanog in F9 cells and observed more than 80% decrease in their expression levels at day3 (Figure 5F). In order to correlate the increase of Sox17 expression with the loss of Oct4 expression, we induced the cells to differentiate into PrE at day 1 after knock-down. "

5. Oct4 is not expressed in XEN cells or primitive endoderm of the embryo. Is Sox17 binding pattern is different in XEN cells than in EC cells at early stages of endoderm differentiation?

Oct4 is required and even sufficient to induce endoderm differentiation (Niwa et al., 2000), but is not required for maintenance of PrE as Referee #2 points out. Mouse XEN cells are well established embryo-derived PrE cells but they express Oct4 at very low level. Therefore we have not looked at Sox17 binding profile in XEN cells by ChIP-seq at this time preferring to concentrate on the F9 system in order to unravel the role of Sox17 to induce endoderm by cooperating with Oct4. However to reconcile Referee #3 and Sox17 ChIP-ChIP data from Niakan et al., 2010 study being available we compared this data with our F9+RA Sox17 ChIP-seq data. We now modify the text p16: "We compared Sox17 ChIP-seq data obtained from F9 +RA cells with Sox17 ChIP-chip data performed in XEN cells which are embryo-derived PrE cells. 1079 genes were found to intersect between the two studies (Figure S4), but only 37 out of these 1079 genes harbour a compressed motif (Table S6), suggesting that these sites are only bound by Sox17 and not by Oct4. This is not surprising as Oct4 is expressed at low levels in XEN cells compared to F9 cells induced to differentiate. These data clearly show that Oct4 and Sox17 co-recruitment on the compressed motif is important for the induction of PrE cell fate, but is likely unnecessary for continued maintenance of PrE." Moreover, Oct4 is expressed in the early PrE of the embryo as demonstrated by numerous studies (Palmieri et

al., 1994 ; Kurimoto et al., 2006 ; Guo et al., 2010) and as we showed in Figure 7B of our manuscript.

6. Figure 6c. The authors state that Sox17 is necessary to recruit Oct4 in RA differentiation. To be able to state this authors need to demonstrate by SiRNA against Sox17 that Oct4 RA peak is lost.

We agree with Referee #2 that generating Oct4 ChIP-seq data in cells knocked-down for Sox17 would reinforce our conclusion that Sox17 is necessary to recruit Oct4 in RA differentiation. However we do not have this data as this is technically very challenging. Indeed, as shown in Figure 8B, Sox7 is also able to cooperatively bind with Oct4 on the compressed motif which can make the data difficult to interpret. We therefore decided to restrain our conclusion and modify the text p16: "Read pile-up analysis revealed that Oct4 is only marginally enriched on compressed motifs before RA induction suggesting that endoderm differentiation induces Sox17 and Oct4 co-recruitment to the compressed motif (Figure 6C)."

7. In Figure 8c authors should also provide reporter expression of enhancers containing canonical motif as a negative control. Flow cyitometry quantification before and after differentiation is also required.

We added the Sox2 canonical enhancer obtained from our Oct4 and Sox2 ChIP-seq data generated from KH2 ESCs expressing Sox2 as a negative control for this experiment in Figure 8C. We also quantified GFP expression by flow cytometry in undifferentiated F9 cells and F9+RA cells induced to differentiate into PrE; results are presented now in Figure S6.

Minor points:

1. Is the data in Figure 1C normalized to the untreated control? What is the fold degree of overexpression of Sox2 and Sox17 in the generated lines?

Figure 1C is normalized to the non-treated control, as specified now in the new version of the figure. Expression levels of exogenous Sox2 and Sox17 relative to the endogenous expression level has been done and is now presented in Supplementary Figure 1.

2. Figure legend 2K is missing and the following ones are shifted up.

The legend for Figure 2K has been added and the following ones annotated in order.

Referee #3:

The authors provide an elegant and important manuscript elucidating how alternative partnering of Oct4 with Sox2 or Sox17 regulates endodermal specification from pluripotent cells. The authors identify and prove the functional importance of a new "compressed" enhancer motif that is targeted by Oct4/Sox17 complexes. The authors also provide conclusive Sox17 mutation swap analysis with Sox2 that is impressive. They also used an embryonal carcinoma model to identify, based on the compressed motif, novel genes up regulated upon endodermal differentiation, and show that their up regulation is dependent both on the presence of Oct4 and Sox17 during this process.

The manuscript is well written with appropriate references. The conclusions are well supported by

the data. I have the following minor points:

1) Throughout the manuscript, the authors use pompous language regarding how they prove the "enhancer code". The manuscript would benefit from toning down the novelty and focus on clarifying the conclusions.

As suggested by Referee #3 we modified the writing of the manuscript throughout when necessary.

a) Abstract: "This work provides new insights in understanding cell fate transcriptional regulation by highlighting the direct link between the DNA sequence and a developmental outcome."

b) p5: "This work helps in our understanding of cell fate transcriptional regulation and suggests a new partner code for Sox/Oct, whereby the recruitment of Sox17/Oct4 to a compressed Sox/Oct motif specifies the endoderm fate."

c) p8: Given these different motif preferences, and our aim to decipher transcriptional regulation that set functionally different gene sets apart, we focused further analysis on Sox2/Oct4 bound canonical and Sox17/Oct4 bound compressed sites.

2) It would be elegant to include Sox2 staining in vivo in early and late ICMs, relative to sox17 (shown in Figure 7B).

We added Sox2 staining to Figure 7B and show that in late blastocysts Sox2 expression is restricted to the epiblast in contrast to Oct4. We now modify the text p17: "This mRNA expression pattern is consistent with Oct4, Sox2 and Sox17 protein levels, with Oct4 uniformly expressed in all cells of the ICM, while Sox2 is initially expressed in all cells of the ICM but then becomes restricted only to the epiblast and Sox17 being initially expressed in a mixed sub-population of ICM cells that later becomes restricted to the blastocoel surface of the ICM (Figure 7B)."

3) The authors should consider revising or omitting their single cells reanalysis presented in Figure 7a. It appears that both extremes of phenotype (Epi-like and Pre-like), have similar Nanog, Sox2 and Gata4 expression levels? I do not understand how they can classify the cells into two groups distinctly based on the expression patterns and curves shown.

We understand Referee #3 concerns but think that figure 7A helps in visualizing that high Sox17 expression correlates with high Oct4 expression at the single cell level in embryos. These data have been collected from embryos at 64 cell stage, and at this stage the epiblast and the primitive endoderm lineages are not yet established. The embryo consists of an ICM with cells acquiring their PrE or EPI identity. Therefore, even if cell specification is en route, it is not surprising to see that some cells seem to be in an intermediate state and show simultaneous expression of PrE and EPI markers like those described. This can also be explained by the fact that we classified the cells into PrE-like or EPI-like based on Sox17 expression as we are interested in Sox17/Oct4 interaction. However, to reconcile Referee #3 suggestion to revise Figure 7A we decided to remove the data concerning Nanog and Gata4 expression. We now modify the text p17: "We observed that the expression pattern of Sox2 is inversely correlated to Sox17 looking across all cells. However, the expression level of Oct4 remained the same in all cells regardless whether they are PrE-like or EPI-like."