Dynamic changes in Sox2 spatio-temporal expression direct the second cell fate decision through Fgf4/Fgfr2 signaling in preimplantation mouse embryos

Tapan Kumar Mistri¹,²,³, Wibowo Arindrarto², Wei Ping Ng¹,², Choayang Wang², Leng Hiong Lim², Lili Sun², Ian Chambers³, Thorsten Wohland¹,⁴,⁵, Paul Robson²,⁶,*

¹Department of Chemistry, National University of Singapore, Singapore, ²Developmental Cellomics Laboratory, Genome Institute of Singapore, Singapore. ³MRC Centre for Regenerative Medicine, Institute for Stem Cell Research, School of Biological Sciences, University of Edinburgh, Edinburgh EH16 4UU, Scotland, United Kingdom, ⁴Department of Biological Sciences and ⁵Centre for Bioimaging Sciences, National University of Singapore, Singapore, ⁶the Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT, USA 06032.

Running Head: Dynamic changes in Sox2 levels direct PE/EPI cell fate decision.
Keywords: Oct4, Sox2, cooperativity, Sox/Oct motif, fluorescence correlation spectroscopy (FCS).

Final character count = 58,469

Address correspondence to:
¹Thorsten Wohland, Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117557. Fax: (65)-6776-7882. Email: twohland@nus.edu.sg
²Paul Robson, The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT, USA 06032. Tel. 207-288-6594 Email: paul.robson@jax.org
ABSTRACT

Oct4 and Sox2 regulate the expression of target genes such as Nanog, Fgf4 and Utf1, by binding to their respective regulatory motifs. Their functional cooperation is reflected in their ability to heterodimerise on adjacent cis regulatory elements, the composite Sox/Oct motif. Given that Oct4 and Sox2 regulate many developmental genes, a quantitative analysis of their synergistic action on different Sox/Oct motifs would yield valuable insights into the mechanisms of early embryonic development. In this study, we measured binding affinities of Oct4 and Sox2 to different Sox/Oct motifs using fluorescence correlation spectroscopy (FCS). We found that the synergistic binding interaction is driven mainly by the level of Sox2 in the case of the Fgf4’s Sox/Oct motif. Taking into account that Sox2 expression levels fluctuate more than Oct4, our finding could explain how Sox2 controls the segregation of the cells of the inner cell mass (ICM) into the epiblast (EPI) and the primitive endoderm (PE) populations within the developing rodent blastocyst.
INTRODUCTION

The mouse preimplantation embryo is a widely used mammalian model to study cell differentiation. Two of the earliest cell fate decisions in mammalian development take place in the preimplantation embryo. The first decision occurs at the 16-32 cell stage and sets apart the morula into two distinct lineages: the trophectoderm (TE) lineage and the inner cell mass (ICM) lineage. At this stage, the TE lineage is a single layer of epithelial cells enclosing the early blastocyst. The ICM lies at one end of the blastocyst, consisting of a pool of pluripotent cells. Later, after embryonic day 3.5 (E3.5), the ICM is further specified into the primitive endoderm (PE) and epiblast (EPI) lineages (Rossant and Tam 2009). Cells of the PE lineage subsequently differentiate into the extra-embryonic cells responsible for secreting patterning cues to the developing embryo proper which consists of cells entirely from the EPI lineage.

The EPI is exclusively characterised by its Nanog, Sox2 and Oct4 expression (Chambers, Colby et al. 2003; Mitsui, Tokuzawa et al. 2003; Silva, Nichols et al. 2009) while the PE is specifically characterised by Gata4, Gata6 and Sox17 (Morrisey, Ip et al. 1996; Morrisey, Tang et al. 1998; Koutsourakis, Langeveld et al. 1999; Aksoy, Jauch et al. 2013). Prior to the segregation into the PE and the EPI, the ICM shows a mosaic pattern of cells expressing Nanog and Gata6 (Chazaud, Yamanaka et al. 2006). The mosaic expression of these markers does not indicate lineage specification as cells expressing the PE markers Gata6 and Gata4, can be coaxed into forming the EPI lineage. The cells only become restricted to their definitive lineages at E4.5 (Chazaud, Yamanaka et al. 2006). However, studies have also shown that inner cells, which have higher Nanog and lower Gata6 expression, give rise to the EPI while cells with lower levels of Nanog and higher levels of Gata6 give rise to the PE (Schrode, Saiz et al. 2014; Xenopoulos, Kang et al. 2015). Therefore, it is not clear what role this difference in expression levels of lineage markers plays in the second cell fate decision of preimplantation development. In addition, how this heterogeneity emerges in the first place has also remained elusive. Studies have indicated that the Fgf4/Fgfr2 signalling pathways lie upstream of this differential expression (Nichols, Zevnik et al. 1998; Guo, Huss et al. 2010; Krawchuk, Honma-Yamanaka et al. 2013). Indeed, Fgf4 is expressed in the EPI lineage but not in the PE while Fgfr2 is expressed in the PE but not in the EPI (Orr-Urtreger, Givol et al. 1991; Niswander and Martin 1992). Treatment with an Fgf signalling inhibitor causes the otherwise mosaic pattern of the ICM cells to generate exclusively the EPI lineage (Yamanaka, Lanner et al. 2011). Furthermore, both Fgf4-null and
**Dynamic changes in Sox2 levels direct PE/EPI cell fate decision**

Mistri et al. 2016

Fgfr2-null embryos are lethal (Feldman, Poueymirou et al. 1995; Arman, Haffner-Krausz et al. 1998). It has been further confirmed that Fgf4 drives, and is required for the segregation of the ICM into the PE and the EPI lineages (Guo, Huss et al. 2010; Kang, Piliszek et al. 2013; Ohnishi, Huber et al. 2014). Therefore, it would be interesting to investigate what controls Fgf4/Fgfr2 signalling, pivotal for the second cell fate decision. Furthermore, several studies showed that spatio-temporal differences in inner cell formation could contribute to the establishment of the heterogeneity in the ICM (Pedersen 1986; Fleming 1987; Krupa, Mazur et al. 2014). Taking these findings into account, it would be of interest to investigate the possible determinants which control FGF/Erk signalling.

In light of the current evidence from mouse preimplantation development, Sox2 emerges as a particularly interesting transcription factor to study. Along with Oct4, it has been found to regulate the expression of other genes important for preimplantation development such as Nanog, Fgf4, Utf1, POU5f1 and Sox2 itself (Nishimoto, Fukushima et al. 1999; Ambrosetti, Schöler et al. 2000; Chew, Loh et al. 2005; Rodda, Chew et al. 2005). In the enhancers of these genes, a Sox2 binding motif, CTTTG(A/T)(A/T) (Harley, Lovell-Badge et al. 1994; Wilson and Koopman 2002) is found adjacent to an octamer motif, ATGC(A/T)AA(T/A) (Verrijzer, Alkema et al. 1992) with a spacer having zero to three base pairs in between the two motifs. Crystallography studies have shown that the Sox2 and Oct4 DNA binding domains heterodimerise on this motif (Reményi, Lins et al. 2003). However, unlike Oct4, Sox2 levels show a dynamic pattern in the preimplantation embryo, in particular, the upregulation specific to the inner cells (Guo, Huss et al. 2010). Additionally, Sox2 is known to be an activator of Fgf4 and a repressor of Fgfr2 (Masui, Nakatake et al. 2007; Chen, Xu et al. 2008). Importantly, a critical level of Sox2 is required for normal development as Sox2-null embryos fail to develop further than immediately after implantation (Avilion, Nicolis et al. 2003). A recent study also highlighted the importance of Sox2 in controlling the expression of Fgf4 (Wicklow, Blij et al. 2014). Collectively, these observations indicate that understanding Sox2 dynamics quantitatively is paramount to understanding whether it plays a role in the cell fate decisions in the preimplantation embryo.

However, despite being controlled by a Sox/Oct motif, the expression levels of Fgf4 and Nanog vary during preimplantation development (Guo, Huss et al. 2010). For example, in the 32-cell stage trophectoderm Fgf4 expression is almost undetectable while Nanog is expressed (Guo, Huss et al. 2010). A possible explanation for this is that the regulatory effect
exerted by Sox2 depends on its expression level. Indeed, Sox2 transcript levels have been shown to fluctuate in a similar fashion to that of *Fgf4* but not with that of *Nanog* (Guo, Huss et al. 2010). Recently, Wicklow et al. showed that Sox2 is necessary for promoting PE differentiation from the ICM via FGF signalling (Wicklow, Blij et al. 2014). Therefore, Sox2 levels might be controlling the cell fate decision of the ICM. However, it is still unclear why *Fgf4* expression is lower than that of Nanog when Sox2 level is low in the early blastocyst and whether Fgf4 expression depends on the critical level of Sox2.

Therefore, we hypothesise that the sensitivity of *Fgf4* and *Fgfr2* expression to Sox2 levels determines cell fate decisions on the part of the ICM. To test this hypothesis, we first investigated the dynamic changes in Sox2 levels in preimplantation development. We studied the direct binding interaction between full-length fusion Sox2 with the *Fgf4 Sox/Oct* motif and the *Fgfr2 Sox* motif, finding their respective apparent dissociations constants (aKd), and investigated the influence of Sox2 level in stable heterodimer formation on the *Fgf4 Sox/Oct* motif compared to on the Sox/Oct motifs of *Nanog* and *Utf1*. We also tested the influence of different bases in the Sox/Oct motifs of known target genes on its enhancer ability. Fluorescent correlation spectroscopy, a single molecule sensitive fluorescence-based technique (Elson and Magde 1974; Jameson, Ross et al. 2009) was used to measure aKds. Remarkably, our results reveal that the formation of a stable Sox2-Oct4-DNA complex on the *Fgf4 Sox/Oct* motif is more dependent on the level of Sox2 than on that of Oct4. Intriguingly, the *Nanog Sox/Oct cis* motif does not show such a high dependency on the level of Sox2 compared to that of the *Fgf4 Sox/Oct cis* motif. This further suggests that dynamic changes in Sox2 levels in the mouse preimplantation embryo could play an important role in the second cell fate decision.
RESULTS

Sox2 level increases in the ICM with time during preimplantation embryo development-

Previously, it has been shown that Sox2 mRNA levels fluctuate more widely than those of Oct4 during preimplantation development (Guo, Huss et al. 2010). In order to test whether Sox2 fluctuations are also present at the protein level, we measured paternally derived zygotic GFP expression from the Sox2 locus in Sox2-null embryos (Ellis, Fagan et al. 2004) following progressive cell stages during preimplantation development (Fig 1a). Furthermore, we decided to reveal the total Sox2 level by immunostaining (Fig 1b). We noticed that the Sox2 level remains ~8 fold higher at the 4-cell stage, ~5 fold higher at the 16-cell stage and again ~8 fold higher at the ICM stage than that of the TE stage (Fig 1c). We observed that Sox2 concentrations reach a maximum in ICM and a minimum in TE at the 64-cell stage which confirmed the notion that the level of Sox2 protein correlates with its mRNA level as had already been hypothesised (Guo, Huss et al. 2010). Whether this observed fluctuation in Sox2 level could play an important role in the regulation of Sox/Oct motif targeted genes is an important question to ask.

Characterization of the Sox/Oct motif sequences-

We first proceeded to characterize the Sox/Oct motif sites using published ChIP-Seq data (Chen, Xu et al. 2008). We created web logos for the Sox/Oct motifs of Oct4 and Sox2 to illustrate these conserved nucleotide positions using the de novo motif identification algorithm CisFinder (Fig. 2a). If an enhancer sequence is conserved across different species in cumulative evolution, the motif is considered to be important for a target gene (Rodda, Chew et al. 2005). Therefore, we further looked through the sequence conservation of the Sox/Oct motif sequence especially for Nanog and Fgf4 across different mammalian species (Fig. 2b). Notably, we observed that the heptamer and octamer sequences of the Sox/Oct motifs irrespective of whether they are Nanog- or Fgf4- specific are highly conserved among the different species analysed. In addition, Rodda et al. showed that the motif remained invariant in the Nanog promoter over an accumulated 250 million years of evolution (Rodda, Chew et al. 2005). This conservation strongly suggests that the Sox/Oct motif requires a conserved sequence to display its function to its target genes. Furthermore, we also observed the conservation pattern of different Sox/Oct motifs from the known genes, Nanog, Utf1, Oct4, Sox2, and Fgf4. A sequence alignment is shown to illustrate the difference in different known Sox/Oct motifs (Fig. 2c).
We noticed that the 2\textsuperscript{nd} and 7\textsuperscript{th} positions in the Sox2 binding site and the 5\textsuperscript{th} - 8\textsuperscript{th} positions in the Oct4 binding site are variable positions among these Sox/Oct motifs.

We further looked for experimental validation of these variable positions in transcriptional activity by luciferase assay. Using the Nanog Sox/Oct \textit{cis} element (NOS) as the reference, we found that mutations in the non-conserved regions of the Sox2 motif affected the transcriptional output in the luciferase assay (Fig. 2d&e). We included mSox, mOct, and mOctSox mutated sequences used by Rodda et al. as reference motifs alongside our mutations NOS2, NOS6, NOS7 and NOS11(Rodda, Chew et al. 2005). Out of these four mutations, NOS2 and NOS7 were applied in variable positions in the Sox binding site, whereas NOS6 and NOS11 were applied in conserved positions of the Sox/Oct motif sequence. We observed a similar trend of transcriptional output from mSox, mOct, and mOctSox as seen previously (Rodda, Chew et al. 2005). In addition, we also noticed a drastic drop in luciferase activity for NOS11 (Fig. 2e). However, NOS6 did not show a similar drop in transcriptional output. This suggests that the 11\textsuperscript{th} position is more important than the 6\textsuperscript{th} in terms of transcriptional activity, although both positions are conserved in the Sox/Oct motif sequence. In contrast, we observed a higher drop in luciferase activity for NOS2 than for NOS7 despite both positions being non-conserved in the Sox/Oct motif sequence. This further illustrates that the 2\textsuperscript{nd} position is favoured in transcriptional activity over the 7\textsuperscript{th}. Therefore, these results suggest that variable positions in the Sox/Oct motif play an important role in gene regulation in addition to the conserved positions of the motif.

Characterization of TFs-fluorescent fusion proteins- To enable measurements by Förster Resonance Energy Transfer (FRET), FCS and EMSA, expression constructs were designed to produce full-length mouse Oct4 and Sox2 fused, via a four amino acid linker (GGSG), with GFP and mCherry, respectively. Initially we tested functionality of both N-terminal and C-terminal fusions. Transient transfections into mouse ES and CHO cells indicated expression of these synthetic constructs and that the resulting fusion proteins were capable of nuclear localization (Fig. 3a). Western blots, using antibodies against the respective transcription factors, in nuclear lysates from ES cells transfected with these constructs further confirmed the expression of the fusion proteins (Fig. 3b). GFP-Oct4 (N-terminal) and mCherry-Sox2 appeared to be expressed at a higher level than their C-terminal tagged counterparts. Importantly, there were only two types of bands detected, namely for the fusion protein
(upper bands close to 75 kDa) and the endogenous protein (lower bands in between 37 to 50 kDa).

We next tested our fusion constructs for their ability to activate transcription in a luciferase promoter assay. CHO cells were used in these experiments as they lack endogenous Oct4 and Sox2 (Fig. S1a). In transient transfection assays, co-transfection of the Oct4 and Sox2 fusion constructs with the wild type Nanog promoter (Wt) resulted in significant luciferase activity above that of the control that lacked the promoter and the mutated promoter (O/S) where mutations applied to the Sox/Oct motif (Fig. 3c), indicating that the fusion constructs have the ability to drive transcription from the Nanog promoter.

We further sought to determine whether these modified TFs retained the ability to bind DNA similarly to their endogenous counterparts. We tested ES cell nuclear lysates transfected with different plasmids contracts necessary for required fusion proteins and incubated with Cy5-labeled DNA containing the Nanog Sox/Oct motif by EMSA. We observed that both endogenous proteins and their fusion counterparts are capable of binding this DNA element as monomers as well as heterodimers (Fig. 3d). Differentially shifted bands indicate that the fusion proteins can heterodimerise with both their respective endogenous and fusion protein partners. The fusion-containing complexes are readily detectable despite these proteins being expressed at lower levels (Fig. 3b) and requiring to compete with their endogenous counterparts in this assay. These results indicate that these fusion proteins are as competent as their wild type proteins in binding to DNA containing the Sox/Oct motif. Together these results add biochemical detail to our finding that fusion Oct4 and Sox2 were as competent as wild type Oct4 and Sox2. Taking into account the higher expression levels for the N-terminal fusion plasmid constructs, we further characterised GFP-Oct4 and mCherry-Sox2 for their physiological functional ability in rescuing ES cells in which the corresponding endogenous TF alleles had been deleted specifically during trophectodermal differentiation (Mistri, Devasia et al. 2015).

The Sox2-Oct4 protein-protein interaction requires DNA - Having demonstrated that the Oct4 and Sox2 fusion proteins perform functionally similarly to their endogenous counterparts, we next sought to utilize GFP-Oct4 and mCherry-Sox2 to quantify their combinatorial binding interplay on the Sox/Oct motif.
Dynamic changes in Sox2 levels direct PE/EPI cell fate decision

Mistri et al. 2016

Using FRET we quantitatively investigated the formation of the Sox2 and Oct4 heterodimer complex with the Nanog Sox/Oct motif in solution using nuclear extracts from transfected CHO cells. We examined Sox2-Oct4 interactions in the presence and absence of DNA to understand the DNA dependency of Sox2-Oct4 complex formation. We did not observe any FRET signal from a solution containing GFP-Oct4 and mCherry-Sox2 in the absence of DNA (Fig. 3e). However, we did notice a distinct FRET signal from the solution containing GFP-Oct4, mCherry-Sox2 and DNA containing Nanog Sox/Oct motif (Fig. 3f). This observation indicates that the DNA brings GFP-Oct4 and mCherry-Sox2 into a close proximity which enables a successful energy transfer from GFP to mCherry, through binding to the Sox/Oct motif.

To further validate DNA-mediated Oct4-Sox2 interactions, we performed FP-EMSA using GFP-Oct4 and mCherry-Sox2 with unlabelled Nanog Sox/Oct motif in CHO nuclear lysate. We only revealed the GFP-Oct4 signal. Again, there was no evidence of an interaction between GFP-Oct4 and mCherry-Sox2 without DNA (Fig. 3g) nor are any multimers of GFP-Oct4 detected. These results indicate that heterodimer complexes are only possible in the presence of DNA which is in agreement with our previous work (Chen, Tapan et al. 2012; Mistri, Devasia et al. 2015).

**Determination of apparent dissociation constants (aKd) by FCS and EMSA**

- Understanding Fgf4 gene regulation requires the quantitative measurement of mCherry-Sox2 binding to the Fgf4 Sox/Oct motif compared to that with the Nanog and Utf1 motifs. For these quantitative studies we used FCS and EMSA as complementary methods. While FCS has been used to measure aKd's in lysate as well as in live cells and zebrafish embryos (Shi, Foo et al. 2009; Mistri, Devasia et al. 2015), FP-EMSA was applied for quantitative aKd measurements for full-length fusion proteins as one can measure the concentration of fusion protein by FCS even in unpurified nuclear lysate (Mistri, Devasia et al. 2015). The titration strategy is shown in Fig. 4a. Direct evidence of complex formation by Fgf4 on the Sox/Oct motif under different titration conditions was compared between EMSA generated gel images and FCS generated ACF curves (Fig. 4b). In the presence of 72 nM mCherry-Sox2, a titration of GFP-Oct4 to the Fgf4 Sox/Oct motif yielded aKd's of 25.2 ± 4.1 nM and 25.3 ± 2.2 nM from EMSA and FCS respectively. (Fig. 4c). On the other hand, a titration of mChery-Sox2 to the same Fgf4 Sox/Oct motif in presence of 40 nM GFP-Oct4 produced aKd of 23.2 ± 1.2 nM and aKd of 24.0 ± 3.0 nM from EMSA and FCS respectively (Fig. 4d). Notably, both
FCS and EMSA provided the similar values within the margins of standard deviation, strengthening the reliability of our quantitative findings on TF-DNA binding interactions.

**Influence of sequence variation in different Sox/Oct motifs on the protein-DNA binding affinity** - Previously, we observed that changes in the conserved and non-conserved regions of the Sox2 motif had an impact on the luciferase activity. Such activity could be linked to the degree of Oct4 and Sox2 binding affinities for these DNA elements and consequently impact the formation of a stable heterodimer. The change in the sequence of the Sox2 binding site “CATTGT\_A” in *Nanog* to “CATTGT\_T” in *Utf1* revealed a slight decrease in the Sox2 binding affinity as the aK\(_d\) increased to 44.0 ± 9.8 nM from 31.7 ± 4.6 nM as measured by FCS. The change in the 7\(^{th}\) position of the sequence “CATTGT\_A” in *Nanog* to “CATTGTG” in *Sox2* revealed a slight decrease in the Sox2 binding affinity as the aK\(_d\) increased to 66.1 ± 18.2 nM from 44.0 ± 9.8 nM. Additionally we observed a slight decrease in the Sox2 binding affinity when both 2\(^{nd}\) and 7\(^{th}\) positions were changed to “CTTTGTG” in *Fgf4* from “CATTGT\_A” in *Nanog* corresponding to an increase to ~ 70 nM for the aK\(_d\) (Fig S3 and Table 1). Our result demonstrates that the variable positions in the heptamer sequence play an important role in the binding interactions of Sox2 with Sox/Oct motif while the conserved positions are the key for strong interactions. The variable positions in the Sox2 binding sequence (CtTTGTt) of different Sox/Oct motifs create diversity in DNA binding affinities of Sox2.

When we consider the Oct4 motif in terms of its binding affinity, we observed that the first four base pairs, ATGC, in the Octamer motif are conserved throughout the five genes. From our luciferase assay (Fig. 2c) we know that this conserved region has an important role in Oct4 binding to its motif; but the non-conserved region (5\(^{th}\) to 8\(^{th}\) position in the octamer motif) also has an important role in the degree of binding affinity as we have seen from the aK\(_d\) values determined on different Sox/Oct motifs. Comparing between the *Oct4* and the *Sox2* Sox/Oct motifs, we found that the 7\(^{th}\) position plays a role in increasing the aK\(_d\) values from 7.7 ± 1.1 nM to 15.9 ± 1.6 nM. We also looked into the *Utf1* and *Fgf4* motifs and observed that the 7\(^{th}\) position plays an important role in increasing the aK\(_d\) value from 32.0 ± 5.5 nM to 42.5 ± 5.5 nM. The 5\(^{th}\) and 6\(^{th}\) position displayed a dramatic change when “AT” was replaced by “TA”; the value increased to 25 nM, as compared to the *Sox2* and *Fgf4* Sox2 binding site (Fig S3 and Table 1). Therefore, our comparative aK\(_d\) measurements
demonstrate the role of variable positions of different Sox/Oct motifs in influencing binding affinity.

Interestingly, we observed that ATGC is highly conserved in the cumulative revolution and therefore, it could be anticipated that ATGC has more influence on Oct4 binding specificity to the octamer sequence of Sox/Oct motifs. Therefore, in addition to these FCS affinity measurements, we performed a FP-EMSA with GFP-Oct4 on differently mutated motifs (see supplementary information) where we applied mutations in the octamer motif sequence (ATGCAAAA). The results showed that Oct4 binding affinity is more strongly correlated with the first 4 base pairs of the octamer motif sequence (ATGC) rather than the last 4 base pairs (AAAA) (Fig. S4a). We further tried to understand the effective influence of a single base pair compared to the collective influence from the conserved base pairs of the octamer sequence (ATGC). In our FP-EMSA experiment, we noticed that the sequence AT has a stronger influence than GC (Fig. S4b). These results indicate that Sox/Oct motif with a mutation in the ATGC region would be less potent in binding with Oct4.

**The role of Sox2 concentration on its synergistic interaction with Oct4** - Having now established that DNA is necessary for the formation of a stable complex between GFP-Oct4 and mCherry-Sox2, we further investigated the importance of protein level for the formation of a stable ternary complex. We measured the *in vitro* $aK_d$ for mCherry-Sox2 and GFP-Oct4 independently and when in solution together by FCS in two separate titrations (Fig. 4a). The main objective was to evaluate whether the ternary complex on the Fgf4 Sox/Oct motif shows any significant response to the level of mCherry-Sox2. We noticed that in the presence of mCherry-Sox2 as a cofactor, GFP-Oct4 showed a higher affinity for the DNA, thus driving exclusive heterodimer formation. Similarly, the presence of GFP-Oct4 as a cofactor aided the binding of Sox2 to these Sox/Oct motifs thus providing evidence that Oct4 and Sox2 have synergistic effects for Nanog, Fgf4 and Utf1 (Table 1). However, we noticed a significant difference in stable ternary complex recruitment among the Fgf4, Nanog and Utf1 Sox/Oct motifs whereby the Fgf4 Sox/Oct motif needs higher levels of Sox2 rather than of Oct4 for the formation of a stable Oct4-Sox2-DNA complex.

The enhanced binding of Oct4 to Nanog, Fgf4 and Utf1 depends on the concentration of Sox2 as validated by the increase in the apparent cooperativity factor at the higher Sox2 cofactor concentration (Table 1). The individual titration of the Fgf4 Sox/Oct motif with
Sox2 gave an $aK_d$ of 70.2 ± 19.1 nM (Table 1). Due to the lower binding affinity of Sox2 to the $Fgf4$ motif as compared to that of Nanog as well as to that of $Utf1$, the influence of Sox2 on Oct4 binding is smaller, giving an apparent cooperativity factor close to 1 at lower Sox2 concentration of 40 nM. At a higher Sox2 concentration of 72 nM, a greater apparent cooperativity factor of 1.7 ± 0.4 was obtained (Table 1). In contrast, Nanog and $Utf1$ showed a similar kind of synergistic effect at a lower Sox2 concentration of 32 nM. This suggests that a higher Sox2 concentration is essential for increasing the binding affinity of Oct4 for the $Fgf4$ Sox/Oct motif. Therefore, from our in vitro titration data we conclude that $Fgf4$ motif needs high levels of Sox2. It has also been reported that Sox2 and $Fgf4$ correlate with each other at the mRNA level during the preimplantation development (Guo et al. 2010). Consistent with this observation, we too found that the protein levels of Sox2 showed similar trend as seen earlier (Fig. 1). Therefore, it will be interesting to address further whether the expression of $Fgf4$ depends on the level of Sox2 in vivo. We sought the answer in Sox2-null embryos by RT-qPCR.

Comparison of Sox2 binding affinity between $Fgf4$ and $Fgfr2$ cis regulatory motifs - It has been previously hypothesized that the $Fgfr2$ gene is a direct target for Sox2 (Chen, Xu et al. 2008). We analysed the published ChIP-Seq data for Sox2 in ES cells and identified a Sox cis motif (Fig. 5a and Fig. S2). We further looked for the sequence conservation of the novel Sox cis motif which illustrated that the motif is only conserved in rodents but with one base pair mismatch in canines, bovines, elephants and opossums. However, it is widely variable for humans. This suggests that this element is important for rodents but not for humans (Roode, Blair et al. 2012). We next sought to verify whether Sox2 directly binds to the novel motif as it does to the $Fgf4$ Sox/Oct motif, as Sox2 was described to be a regulator of earlier $Fgf4$ and $Fgfr2$ genes expression (Guo, Huss et al. 2010) Furthermore, Masui et al., 2007 had experimentally shown that the absence of Sox2 favours $Fgfr2$ upregulation which suggests that Sox2 may be working as a repressor of $Fgfr2$ (Masui, Nakatake et al. 2007). We performed an EMSA assay using mCherry-Sox2 in CHO nuclear cell lysate with the $Fgfr2$ and $Fgf4$ motifs. We noticed that mCherry-Sox2 formed a stable monomer with both DNA elements (Fig. 5b). We further quantified the binding affinities of Sox2 to these motifs by FCS (Fig 5c&d). Our result showed that both $Fgfr2$ ($aK_d$ values is 81.2 ± 15.1 nM) and $Fgf4$ ($aK_d$ values is 70.2 ± 19.1 nM) require high concentrations of Sox2 for stable complex formation (Fig. 5c&d). However, the presence of the Oct4 binding motif in the $Fgf4$ Sox/Oct motif favours stable complex formation even at low concentrations of Sox2 (Table 2).
could be the reason for the good correlation between $Fgf4$ expression and Sox2 level during preimplantation development. On the other hand, $Fgfr2$ shows minimal expression where the Sox2 level is high such as in the EPI and the opposite happens in the TE lineage. This suggests that Sox2 works as an activator of $Fgf4$ and repressor of $Fgfr2$. Therefore, it will be interesting to address further whether the expression of $Fgf4$ depends on the level of Sox2 	extit{in vivo}. We sought the answer in Sox2-null embryos by RT-qPCR.

**Validation of the role of Sox2 on its target genes in Sox2-null embryos by RT-qPCR** -
Taken together, our present and published work (Guo, Huss et al. 2010) suggests that Sox2 works as regulator for both $Fgf4$ (positively) and $Fgfr2$ (negatively). To investigate this further, we performed RT-qPCR in total mRNA samples derived from Sox2-null embryos (Ellis, Fagan et al. 2004). We noticed that the expression of $Fgf4$ is minimal compared to that of the housekeeping gene, Actin $\beta$. On the other hand, expression of $Fgfr2$ is higher than that of Actin $\beta$ (Fig. 6b). We also observed that expression of $Nanog$ is not influenced by the absence of Sox2 which further supports our 	extit{in vitro} data where we observed that the $Nanog$ Sox/Oct motif is capable of making stable ternary complexes at lower Sox2 concentrations than the $Fgf4$ Sox/Oct motif. It should be noted that maternal Sox2 is still present in early blastocyst and this could be sufficient for the expression of $Nanog$ but not for that $Fgf4$ and $Fgfr2$ as they need higher Sox2 levels (Avilion, Nicolis et al. 2003). RNA-Seq data further support our findings as the expression of $Fgf4$’s was minimal in Sox2-null embryo in comparison with wild type embryos (data not shown).
DISCUSSION

Whether the segregation of ICM into PE and EPI is driven by stochastic or deterministic events is currently controversial (Chazaud, Yamanaka et al. 2006; Morris, Teo et al. 2010). However, this differentiation event depends on Fgf/Erk signaling which in turn depends upon communication between inner cells expressing Fgf4 and inner cells expressing Fgfr2 (Chazaud, Yamanaka et al. 2006; Guo, Huss et al. 2010; Kang, Piliszek et al. 2013; Ohnishi, Huber et al. 2014). In this study, we provide evidence in support of a model in which temporal alterations in the Sox2 concentration differentially regulate expression of Fgf4 and Fgfr2, thereby driving segregation of ICM into PE and EPI.

To assess key differences in the protein-DNA binding interactions in the light of what is known about the second differentiation event of mouse embryonic development; we measured the binding kinetics of Oct4 and Sox2 to different Sox/Oct enhancers specific to Nanog, Fgf4 and Utf1. Our in vitro protein-DNA binding analyses indicate that stable protein-DNA complex formation is dependent not only on the DNA sequence specificity but also on the concentration of proteins involved. Interestingly, binding of Sox2 to the Fgf4 Sox/Oct motif requires a higher concentration of Sox2 (~2 fold) than is needed for similar complex formation on Sox/Oct motifs from Nanog or Utf1. Direct binding of Sox2 to the Fgfr2 and Fgf4 Sox/Oct motifs and the requirement of a high Sox2 concentration for formation of a stable Sox2/DNA complex further lead us to question whether the expression of Fgf4/Fgfr2 correlates with the high level of Sox2 in vivo. In Sox2-null embryos at E3.5 (and therefore, in the absence of paternal Sox2), Fgf4 is down-regulated and Fgfr2 is up-regulated. These findings argue that Fgf4/Fgfr2 expression is highly dependent on the Sox2 concentration.

Based on our in vitro and in vivo results, we propose a model considering Fgf4 and Nanog which adds more clarity to the second cell fate decision during mouse development (Fig. 6a). At 37 nM Sox2, the aKd value for the Nanog Sox/Oct motif is 13.8 ± 3.0 nM with an apparent cooperativity factor of 2.0 ± 0.6 (Table 1), while the aKd for the Fgf4 Sox/Oct motif is 42.5 ± 5.5 nM with an apparent cooperativity factor of 0.9 ± 0.1 (Table 1). Therefore, at low Sox2 concentration, such as 37nM, Nanog may be expressed but Fgf4 may not be expressed. This agrees with mRNA expression analysis (Fig. 6b) at E3.5, showing that Nanog is expressed but Fgf4 is not. At 72 nM Sox2, the aKd value and apparent cooperativity
factor for the *Fgf4* Sox/Oct motif was 25.2 ± 4.1 nM and 1.7 ± 0.4 respectively, whereas the aK_d and the apparent cooperativity factor for the *Nanog* motif were relatively unchanged. At this elevated Sox2 concentration both *Nanog* and *Fgf4* may be expressed.

From the above argument, we propose the gene regulation model illustrated (Fig. 6c) controlling the segregation of the ICM into the EPI and the PE. As the zygotic Sox2 expression is detectable in morula, we consider development from the morula to the late blastocyst. The morula, 16-cell stage embryo, consists of a group of inner cells surrounded by outer cells. In the inner cells, the zygotic expression of Sox2 starts but initially at a concentration not high enough to drive upregulation of *Fgf4* and downregulation of *Fgfr2*. In contrast, expression of Sox2 at this stage is absent in the outer cells and the maternal Sox2 protein level is depleting, resulting in downregulation of *Fgf4* and upregulation of *Fgfr2*. This scenario becomes more critical for the outer cells at the end of another round of cell division where the Sox2 level decreases further resulting in no expression of *Fgf4* and *Nanog*; but the clear expression of *Fgfr2*. At the end of the 32 cell stage, those outer cells behave like presumptive trophectoderm (TE). Prior to the cavitation process, there remains heterogeneity among inner cells, due to a second wave of inner cell formation after cell division of the 16-cell morula. At the end of the 32-cell stage, the early inner cells (deep purple) will already be expressing *Fgf4* and downregulate *Fgfr2* whereas the late inner cells (light purple) have high levels of *Fgfr2*. These considerations are consistent with a previous report that also suggested that cells generated in the second wave express higher levels of *Fgfr2* than those from the first wave (Morris, Graham et al. 2013).

After cavitation, the late morula proceeds to the early blastocyst stage where the outer cells are already fated to become TE and the mixed population of early and late inner cells create the ICM. We assigned to the early inner cells in the ICM as the label of presumptive EPI and to the late inner cells that of presumptive PE. At this stage of embryo development, the early inner cells possess a high level of *Fgf4* whereas the late inner cells possess a high level of *Fgfr2*. Therefore, the late inner cells come in direct contact with the signalling output of *Fgf4*-expressing early inner cells, resulting in the downregulation of a number of pluripotency genes including, *Nanog*, *Klf2*, and *Esrrb* in the late inner cells and upregulation of PE-specific genes including *Gata6*, *Gata4*, and *Sox17*. Therefore, the late inner cells go on to become the PE lineage. In contrast, the early inner cells maintain the upregulation of the pluripotency genes and increased zygotic Sox2 levels replenish the depleting level of
maternal Sox2 protein. Consequently, they commit to the EPI lineage. In summary, the \textit{in vitro} protein-DNA binding data and \textit{in vivo} analysis of Sox2 levels controlling the expression of \textit{Fgf4} and \textit{Fgfr2}, allows us to argue that Sox2 increasing levels in the inner cells could be a determinant for the segregation of the ICM into the PE and the EPI cell lineages.

\textbf{MATERIALS AND METHODS}

\textbf{Electrophoretic mobility shift assay (EMSA)-} EMSA was carried out as described (Rodda, Chew et al. 2005; Chen, Tapan et al. 2012). Details of quantitative titration assay are provided in \textit{supplementary information}. Additionally a fluorescence protein based EMSA (FP-EMSA) was also performed as described (Chen, Tapan et al. 2012; Hutchins, Choo et al. 2013). All the oligonucleotides used for EMSA are provided in the \textit{supplementary information}.

\textbf{Luciferase Assay -} CHO cells were cultured in Dulbecco’s modified Eagle’s medium with high glucose (Invitrogen), 10% standard fetal bovine serum (Hyclone), and 1% penicillin/streptomycin and maintained at 37\degree C with 5\% CO\textsubscript{2}. For a 24-well plate, 0.5 \mu g of EO plasmid, 0.5 \mu g of MS plasmid and 0.3 \mu g of pGL3 NSO plasmid were co-transfected per well using Lipofectamine 2000 (Invitrogen) according to the manufacturer’s instructions. 0.05 \mu g of \textit{Renilla} luciferase plasmid (pRL-TK from Promega) was co-transfected as an internal control. Firefly and \textit{Renilla} luciferase activities were measured 24 hours post transfection using the Dual Luciferase Kit (Promega) and a Centro LB960 96-well luminometer (Berthold Technologies). Alternatively, F9 embryonal carcinoma (EC) cells were used in the luciferase assay for motif characterization to understand the importance of variable positions in the Sox/Oct motif sequence. Cell culture, transfection and sample preparation for F9 EC cells were performed as described earlier (Rodda, Chew et al. 2005).

\textbf{Concentration measurement of fusion protein -} The concentration of fusion proteins in unpurified nuclear lysate by fluorescence correlation spectroscopy (FCS) was measured as described (Buschmann, Krämer et al. 2009; Mistri, Devasia et al. 2015). See the \textit{supplementary section} for the theoretical models used in FCS analysis.

\textbf{Data analysis (EMSA& FCS) for aK\textsubscript{d} determination -} The apparent dissociation constants, aK\textsubscript{d}, were determined as described earlier (Mistri, Devasia et al. 2015).
**Immunocytochemical Staining and Image J based semi-quantification** - Embryos were fixed in 2.5% PFA for 15 min at 37°C, washed with Triton (0.1% in PBS; 5 min), Triton (0.5% in PBS; 20 min), Triton (0.1% in PBS; 5 min), and BSA/Tween (0.1% BSA and 0.01% Tween in PBS; 30 min). After incubation with 1° antibody (Sox2-Y17) in BSA/Tween (60 min), embryos were washed with BSA/Tween (3 x 15 min), and then incubated with 2° antibody (Goat anti-Rabbit IgG) conjugated with Alexa Fluor 488 (Molecular Probes, Carlsbad, CA) for an additional hour. Following BSA/Tween washes (3 x 15 min), embryos were passed through increasing concentrations of mounting solution containing To-Pro prior to final mounting. Images were captured with a confocal microscope (LSM 510 META; Zeiss, Thornwood, NJ). All the Z-stack images from an individual embryo were grouped into one stack picture based on average fluorescence intensity employing image J software (NIH). Nuclear staining dye To-Pro was used as a control for normalizing the fluorescence intensity of Sox2 targeted antibody tagged with Alexa Fluor 488. It is known that intensity varies with respect to the absolute concentration and hence the ratio of fluorescence intensity of Sox2 to that of To-Pro will be the equal to the ratio of concentration of Sox2 to that of To-Pro. The normalized Sox2 concentration ratio, which is an absolute parameter, was compared across the different cell stages of early mouse embryo.

**Time laps imaging for detection of GFP signal in growing embryos in culture** - Two cell-stage embryos, derived by crossing of Sox2-GFP heterozygotic females (Ellis, Fagan et al. 2004) and males, were cultured in M2 media on the Zeiss microscope (Axio Observer D1, Zeiss) stage in appropriate culture conditions (CO2-5%, 37°C). Fluorescence images were captured in 6-hour intervals until the last blastocyst stages. GFP photo-bleaching was minimised by using a low laser power (10 µW) from 488 nm laser. The scale bar was 20 µM.

**Real time quantitative polymerase chain reaction (RT-qPCR)** - All the mouse work was approved by the BRC IACUC (Biopolis). Embryos were derived by crossing of Sox2-GFP heterozygotic females (Ellis, Fagan et al. 2004) and males and collected at 3.5 dpc in M2 medium. Total RNA was extracted and purified from the whole embryos using a PicoPure RNA isolation kit (Arcturus Bioscience) and cDNA was synthesized with a high capacity cDNA archive kit (Applied Biosystems; ABI). cDNA was first pre-amplified with a pool of 48 inventoried Taqman assays (20X, Applied Biosystem) by denaturing at 95°C for 15 s, and annealing and amplification at 60°C for 4 min for 14 cycles. The pre-amplified products were...
diluted 5-fold and the expressions of the 48 assays were analyzed with 48/48 Dynamic Arrays on a Biomark System (Fluidigm). Ct values were calculated from the system’s software (Biomark Real-time PCR Analysis, Fluidigm). See the supplementary information for further details of methods and materials part.

ACKNOWLEDGEMENT

We sincerely thank, Dr Stefan Barakat and Mr Arun George Debasia for their valuable suggestions in the manuscript. We are also grateful to Dr David Rodda, Dr Kamesh Narayan and Dr Foo Yong Hwee for their technical assistance in the project. We also like to thank Ms. Elisa Hall-Ponsele for proof reading. TKM gratefully acknowledges support by a National University of Singapore graduate scholarship. Research in the IC lab was supported by the Medical Research Council of the UK. Research in the PR lab was supported by the Agency for Science Technology & Research (A*STAR). Research in the TW lab was supported by a NUS-BW (National University of Singapore / Baden-Württemberg, R-143-000-422-646) joint grant to TW and PR.

AUTHOR CONTRIBUTIONS

T.K.M. did experiments and wrote the manuscript, W.A. helped T.K.M in construct preparation and manuscript preparation, W.P.N helped T.K.M in FCS experiment, L.S. did immunostaining experiment, C.W. performed RT-PCR experiment, I.C. improved the manuscript by re-writing and overall work is supervised by T.W. and P.R.

ADDITIONAL INFORMATION

Please see the Supplementary Section for further information.
Dynamic changes in Sox2 levels direct PE/EPI cell fate decision

Mistri et al. 2016

REFERENCE


Wicklow, E., S. Blij, et al. (2014). "HIPPO pathway members restrict SOX2 to the inner cell mass where it promotes ICM fates in the mouse blastocyst."

FIGURE LEGENDS

FIGURE 1. In vivo localisation and quantification of Sox2 level during mouse pre-implantation development.
(a) Microscopic observation of paternally derived zygotic Sox2-GFP expression in Sox2-null embryos (Guo, Huss et al. 2010) with the progression of time until 114 hrs from single cell zygote (guided by an arrow). GFP expression from the Sox2 locus is observed starting from 66 hrs at 8-cell stage. The middle embryo is a wild type control.
(b) Confocal Z-stack images of representative Sox2 and ToPro nuclear stained embryos.
(e) Quantification of the average Sox2 level per cell in 4-16 cell embryos or in TE or ICM of 64 cell embryos shown in (b).

FIGURE 2. Characterisation of Sox/Oct cis motifs.
(a) The top de novo sequence motifs (based on enrichment) detected by CisFinder (Sharov and Ko 2009) in Oct4 and Sox2 ChIP-Seq data from ESCs (Chen, Xu et al. 2008), shown by WebLogo.
(b) Comparison of Nanog and Fgf4 Sox/Oct cis motifs across different eutherian species. A phylogenetic tree illustrated the sequence conservation of those Sox/Oct motifs in cumulative evolution.
(c) Alignment of the Sox/Oct motifs from the target genes Nanog, Utf1, Oct4, Sox2 and Fgf4.
(d) Mutations generated in the Sox/Oct motif in a 400 bp gene fragment from Nanog located at -289 to +117 (relative to the Nanog transcription start site) (Chambers 2005; Rodda, Chew et al. 2005) were subsequently tested.
(e) Luciferase activity of constructs shown in (d) after transfection of F9 embryonal carcinoma cells

FIGURE 3. Functional determination of synthetic fusion protein constructs.
(a) Confocal images of ES cells and CHO cells transfected with pCAG-GFP-Oct4-IN and pCAG-mCherry-Sox2-IN.
(b) Immunoblot analysis of exogenous fusion proteins; bands A and B correspond to the fusion and wild type proteins respectively.
(c) Dual luciferase assays comparing the transcriptional activity of the wild type Nanog Sox/Oct motif (wt) (CATTGTAATGCAAAA), a mutated Nanog Sox/Oct motif (O/S) (CAGGTTAATTTGAAA), and a Nanog promoter with a deletion form X – y that removes ONLY the Sox/Oct motif (control).
(d) DNA binding of fusion proteins (GFP scan) compared to wild type proteins (Cy5 scan) by EMSA from ES cells nuclear lysates. A – heterodimer complex (GFP-Oct4/mCherry-Sox2/NSO element); B – heterodimer complex (GFP-Oct4/wild type Sox2/NSO element); C – monomer complex (mCherry-Sox2/NSO element); D – monomer complex (GFP-Oct4/NSO element).
(e-f) Emission intensity collected at >500 nm wavelength upon GFP excitation by 488nm laser. FRET analysis was performed on nuclear extracts isolated from transfected CHO cells in the absence (e) or the presence (f) of DNA containing the Nanog Sox/Oct motif. The arrow indicates the emission maxima of mCherry.
(g) Complex formation shown by FP-EMSA detecting GFP-Oct4 in the presence or absence of unlabelled DNA and mCherry-Sox2.

FIGURE 4. Comparison of FCS and EMSA derived aKd.
(a) Conceptual scheme showing the titration of GFP-Oct4 in the absence and presence of mCherry-Sox2 and vice versa.
(b) EMSA scans and autocorrelation curves generated from FCS assays for the Fgf4 Sox/Oct motif titrating with (I) GFP-Oct4; (II) mCherry-Sox2; (III) GFP-Oct4 in the presence of a
Dynamic changes in Sox2 levels direct PE/EPI cell fate decision Mistri et al. 2016

fixed mCherry-Sox2 concentration and (IV) mCherry-Sox2 in the presence of a fixed GFP-Oct4 concentration. A: monomer (GFP-Oct4/DNA); B: monomer (mCherry-Sox2/DNA); C: heterodimer (GFP-Oct4/mCherry-Sox2/DNA); D: free Cy5-DNA. (c-d). Comparison of the bound fraction vs. total protein concentration plots with the Fgf4 Sox/Oct motif obtained by EMSA (left panel) or FCS (right panel).

FIGURE 5. Direct binding of Sox2 protein to cis elements of Fgf4 and Fgfr2.
(a) Identification of the Sox cis regulatory element near the transcription start site of Fgfr2 from Sox2 ChIP-Seq data in ES cells (Chen, Xu et al. 2008). Sequence conservation of the novel motif across different mammalian species is shown.
(b) EMSA comparison of the binding of Sox2 to the novel Sox cis motif and the Fgf4 Sox/Oct cis motif; lane 1: Fgfr2 element, Lane 2: Fgf4 element. Yellow and white stars indicate Sox2-DNA complex and free DNA respectively.
(c-d) Two representative normalized autocorrelation (ACF) curves at no Sox2 as well as at high level of Sox2 shown for Fgfr2 (c) and Fgf4 (d) elements. The measured aKd's from each titration assay measured by FCS technique were compared.

FIGURE 6. Bridging in vitro measurements with in vivo Sox2 levels of early embryos.
(a) Hypothetical diagram showing relation of aKd to gene expression with respect to Sox2 concentration is depicted.
(b) Relative mRNA levels of Fgf4, Nanog, and Fgfr2 in Sox2-null embryos at E3.5 were compared to that of WT embryos by RT-qPCR.
(c) Gene regulation model for Fgf4, Nanog and Fgfr2 during the segregation of ICM into PE and EPI by Bio Tapestry software (Longabaugh 2012). The thickness of the blue lines refers to the Sox2 level; the black ball indicates Oct4. Deep and light purple refer to the early inner cells and the late inner cells respectively. Green shade refers to trophectodermal lineage.

TABLE 1 (a) Synergism correlates with the Sox2 concentration. aKd measurements from different titrations of GFO-Oct4 against Nanog, Fgf4, and Utf1 elements in the presence and absence of mCherry-Sox2 by FCS and vice versa.

Supplementary Fig 1. CHO cell line is ideal for fusion protein generation. (a) Western blot results for Oct4 (left) and Sox2 (right) in ES or CHO nuclear cell lysates. (b) Complex
formation of GFP-Oct4 with the Nanog Sox/Oct motif in the presence or absence of mCherry-Sox2 by FP-EMSA.

Supplementary Fig 2. **Comparison of natural DNA sequences among Fgfr2, Fgf4, and Nanog motifs.** 37 bps long DNA elements containing either of Fgfr2, Fgf4, or Nanog motif were compared for all three oligos (top), Fgfr2 and Fgf4 (middle) and Fgfr2 and Nanog (bottom). Yellow colour represents base pair similarity among variants and blue colour represents base pair similarity for less than 3 variants.

Supplementary Figure 3: **aKₐ of Sox/Oct motifs measured for either Oct4 or Sox2.** The indicated Sox/Oct motifs were compared to show the Sox2 binding site sequence (left) and the Oct4 binding site sequence (right) as well as aKₐ values measured for either Sox2 or Oct4 alone by FCS.

Supplementary Figure 4: **Importance of sequence specificity in Oct4-DNA binding on Nanog Sox/Oct motif.** (a) The binding affinity of Oct4 to the Nanog Sox/Oct motif or to mutations within the octamer (ATGCAAAA) sequence is compared. Controls are shown in lanes 1 to 8 and mutation experiments in lanes 9 to 15. GFP-Oct4 was added in all lanes except lane 7 (untransfected CHO nuclear lysate). (b) The binding affinities of Oct4 to the Nanog Sox/Oct cis motif or to different mutations within the octamer (ATGC) sequence were compared. Controls were shown in lanes 1 to 8 and mutation experiments in lanes 9 to 15. GFP-Oct4 was added in all lanes except lane 7 (untransfected CHO nuclear lysate). All oligonucleotide sequences are in the [supplementary information](#).
Mistri et al. 2015

Figure 1

(a) GFP, Brightfield, Merge

(b) Sox2, ToPro, Brightfield

Post-fertilization hours (in vitro culture)

(c) Intensity of Sox2 transcription in relation to TE and ICM
Sox-Oct cis element from Oct4 ChIP-Seq peaks

Sox-Oct cis element from Sox2 ChIP-seq peaks

Mouse Sox-Oct cis element targets

Tested different mutations in Sox-Oct cis element

Sox-Oct cis element targets

Eutheriana Nanog Sox-Oct cis element

Position of those mammalian animals in cumulative evolution

Eutheriana Fgf4 Sox-Oct cis element

Mistri et al. 2015

Figure 2

http://dx.doi.org/10.1101/052530 (not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license (which was first posted online Jun. 22, 2016; doi: bioRxiv preprint).
**Mistri et al. 2015**

**Figure 3**

In vitro FRET study in the absence of DNA

- eGFP-Oct4 + mCherry-Sox2
- mCherry-eGFP
- mCherry + eGFP

**Graph (c)**

- Relative luciferase activity
- E14 mES cell
- CHO cell

**Graph (e)**

- Wavelength (nm)
- Relative fluorescence intensity
- eGFP-Oct4 + mCherry-Sox2
- mCherry-eGFP
- mCherry + eGFP

**Graph (f)**

- In vitro FRET study in the presence of DNA
- Relative fluorescence intensity
- eGFP-Oct4 + mCherry-Sox2
- mCherry-eGFP
- mCherry + eGFP
Mistri et al. 2015
Figure 4

a) Titration of mCherry-Sox2 with and without Oct4

Titration of GFP-Oct4 with and without Sox2

DNA Motif  eGFP-Oct4  mCherry-Sox2  Cy5  Fixed concentration

b) Titration experiment

<table>
<thead>
<tr>
<th>(Fgf4 element)</th>
<th>EMSA</th>
<th>FCS</th>
<th>Autocorrelation curves</th>
</tr>
</thead>
<tbody>
<tr>
<td>II [mCherry-Sox2]</td>
<td>[mCherry-Sox2] B</td>
<td>[mCherry-Sox2] D</td>
<td>[mCherry-Sox2] D</td>
</tr>
<tr>
<td>IV [mCherry-Sox2] with eGFP-Oct4</td>
<td>[mCherry-Sox2] C</td>
<td>[mCherry-Sox2] A</td>
<td>[mCherry-Sox2] A</td>
</tr>
</tbody>
</table>

Titration of mCherry-Sox2 with and without Oct4

Titration of GFP-Oct4 with and without Sox2

Titration of eGFP-Oct4 with mCherry-Sox2

Titration of mCherry-Sox2 with eGFP-Oct4

<table>
<thead>
<tr>
<th>Bound Fraction</th>
<th>Concentration of eGFP-Oct4 (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>10^10</td>
</tr>
<tr>
<td>0.2</td>
<td>10^9</td>
</tr>
<tr>
<td>0.4</td>
<td>10^8</td>
</tr>
<tr>
<td>0.6</td>
<td>10^7</td>
</tr>
<tr>
<td>0.8</td>
<td>10^6</td>
</tr>
<tr>
<td>1.0</td>
<td>10^5</td>
</tr>
</tbody>
</table>

Concentration of eGFP-Oct4 = 25.3 ± 2.2 nM

Concentration of mCherry-Sox2 = 24.0 ± 3.0 nM

Concentration of eGFP-Oct4 = 25.2 ± 4.1 nM

Concentration of mCherry-Sox2 = 23.3 ± 1.2 nM

FCS

EMSA

DNA Motif  eGFP-Oct4  mCherry-Sox2  Cy5  Fixed concentration

Mistri et al. 2015

FCS

EMSA
Fgfr2 motif

FGF4 motif

Normalized ACF Curves

No Sox2

With Sox2

\( \text{aK}_d = 81.2 \pm 15.1 \text{ nM} \)

\( \text{aK}_d = 70.2 \pm 19.1 \text{ nM} \)

Mistri et al. 2015

Figure 5

http://dx.doi.org/10.1101/052530

CC-BY-NC-ND 4.0 International license

not peer-reviewed) is the author/funder. It is made available under a

The copyright holder for this preprint (which was.
**Comparison between two cis elements**

- Early Morula Expression
- Epiblast Expression

**Concentration of Sox2 (nM)**

- $K_i = 25.2 \pm 4.1$ nM at 70 nM
- $K_i = 13.8 \pm 3.0$ nM at 370 nM

**Bound fraction comparison**

- Nanog

**Fold change in Sox2-null embryo at E3.5**

- $\beta$ Actin
- Fgf4
- Nanog
- Sox2
- Fgf2

**Mistri et al. 2015**

Figure 6

http://dx.doi.org/10.1101/052530

CC-BY-NC-ND 4.0 International license (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.
Table 1: **aK\textsubscript{d} values obtained from titration assays.** The measurement were performed by FCS at constant DNA concentration (5 nM) using unpurified CHO nuclear lysate containing either GFP-Oct4 or mCherry-Sox2. The symbol “*” represents the apparent cooperativity factor.

<table>
<thead>
<tr>
<th>Sox-Oct element</th>
<th>Titration with GFP-Oct4</th>
<th>Titration with mCherry-Sox2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Apparent dissociation constant, aK\textsubscript{d} (nM)</td>
<td>No Sox2</td>
</tr>
<tr>
<td>Nanog element</td>
<td>28.2 ± 4.9</td>
<td>13.8 ± 3.0 (2.0 ± 0.6)*</td>
</tr>
<tr>
<td>Fgf4 element</td>
<td>42.5 ± 5.5</td>
<td>49.1 ± 5.0 (0.9 ± 0.1)*</td>
</tr>
<tr>
<td>Utf1 element</td>
<td>32.0 ± 5.5</td>
<td>19.8 ± 7.3 (1.6 ± 0.7)*</td>
</tr>
</tbody>
</table>
a

Western Blot for Oct4

<table>
<thead>
<tr>
<th></th>
<th>ES cell</th>
<th>CHO cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wild type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>eGFP-Oct4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mCherry-Sox2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wild type</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Western Blot for Sox2

<table>
<thead>
<tr>
<th></th>
<th>ES cell</th>
<th>CHO cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wild type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>eGFP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mCherry-Sox2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wild type</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b

<table>
<thead>
<tr>
<th></th>
<th>eGFP scan</th>
<th>Cy5 scan</th>
<th>Merge</th>
</tr>
</thead>
<tbody>
<tr>
<td>eGFP-Oct4</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>mCherry-Sox2</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Sox-Oct element</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

A

B

A

1 2 3 1 2 3 1 2 3
<table>
<thead>
<tr>
<th>Sox2 binding site</th>
<th>K&lt;sub&gt;d&lt;/sub&gt; (nM)</th>
<th>Oct4 binding site</th>
<th>K&lt;sub&gt;d&lt;/sub&gt; (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nanog</td>
<td>CATTTTTA</td>
<td>12345678</td>
<td>ATGCAAAAA</td>
</tr>
<tr>
<td>Utf1</td>
<td>CATTTTT</td>
<td>12345678</td>
<td>ATGCTAGT</td>
</tr>
<tr>
<td>Oct4</td>
<td>CTTTTTT</td>
<td>12345678</td>
<td>ATGCACTCT</td>
</tr>
<tr>
<td>Sox2</td>
<td>CATTTTG</td>
<td>12345678</td>
<td>ATGCAATA</td>
</tr>
<tr>
<td>Fgf4</td>
<td>CTTTTTT</td>
<td>12345678</td>
<td>ATGCTAAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>31.7 ± 4.6</td>
<td>28.2 ± 4.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>44.0 ± 9.8</td>
<td>32.0 ± 5.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>72.3 ± 9.9</td>
<td>7.7 ± 1.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>66.1 ± 18.2</td>
<td>15.9 ± 1.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>70.2 ± 19.1</td>
<td>42.5 ± 5.5</td>
<td></td>
</tr>
</tbody>
</table>