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

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

Tapan Kumar Mistri[1, 2, 3,4], Wibowo Arindrarto[3], Wei Ping Ng[2,3], Choayang Wang[3], Leng Hiong Lim [3], Lili Sun[3], Ian Chambers[4, +], Thorsten Wohland[2, 5, 6, x], Paul Robson[3, 7,*]

[1]School of Chemical Engineering and Physical Sciences, Lovely Professional University, Phagwara, Punjab (India)-144411,[2]Department of Chemistry, National University of Singapore, Singapore, [3]Developmental Cellomics Laboratory, Genome Institute of Singapore, Singapore. [4]MRC Centre for Regenerative Medicine, Institute for Stem Cell Research, School of Biological Sciences, University of Edinburgh, Edinburgh EH16 4UU, Scotland, United Kingdom,[5]Department of Biological Sciences and [6]Centre for Bioimaging Sciences, National University of Singapore, Singapore, [7]The 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 = 48,355

Address correspondence to:

[+]Ian Chambers, School of Biological Sciences, University of Edinburgh, 5 Little France Drive, Edinburgh EH16 4UU, Scotland. Tel: +44 131 651 9500; E-mail: ichambers@ed.ac.uk

[x]Thorsten Wohland, Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117557. Fax: (65)-6776-7882. Email: twohland@nus.edu.sg

[*]Paul 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 motifs, 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* Sox/Oct motif. Taking into account *Sox2* expression levels fluctuate more than *Oct4*, our finding provides an explanation on how Sox2 controls the segregation of the epiblast (EPI) and primitive endoderm (PE) populations within the inner cell mass (ICM) of 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 trophoblast, represented by the trophectoderm (TE) and the inner cell mass (ICM). At this stage, the TE 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 [1]. Cells of the PE lineage subsequently differentiate into the extra-embryonic cells responsible for secreting patterning cues and providing nutrition to the developing embryo proper which consists of cells entirely from the EPI lineage.

The EPI is exclusively characterised by its *Nanog and Sox2* expression [2-4] while the PE is specifically characterised by *Gata4, Gata6* and *Sox17* [5-8] while Oct4 initially persists in both [8]. Prior to the segregation into the PE and the EPI, the ICM shows a mosaic pattern of cells expressing *Nanog* and *Gata6* [9]. 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 [9]. 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 [10, 11]. 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 pathway lies upstream of this differential expression [12-14]. 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 [15, 16]. The segregation of PE from the EPI is also observed to be dependent on FGF/Erk signalling where the entire bipolar ICM can acquire pluripotency if this signal is absent [9, 17]. Additionally, a treatment with an Fgf signalling inhibitor causes the otherwise mosaic pattern of the ICM cells to generate exclusively the EPI lineage [13, 18]. Recently it is also reported that p38 family mitogen activated protein kinases (p38-Mapk14/11) actively participate in the 2nd cell fate determination; especially during early-blastocyst maturation for assisting bipolar ICM cells.

Interestingly, as like Erk1/2, Fgf-receptor signalling controls the functional activation of p38-Mapk14/11 [19]. Furthermore, both *Fgf4*-null and *Fgfr2*-null embryos are lethal [20, 21]. It has been further confirmed that *Fgf4* is required for the segregation of the ICM into the PE and the EPI lineages [13, 22, 23]. Furthermore, several studies indicate spatio-temporal differences in inner cell formation contribute to the establishment of the heterogeneity in the ICM [24-26]. Recently, Kang et al. showed that Fgf4 is the central molecule for determining the distinct lineages from ICM cells and Fgf4 imparts its action with the help of Fgfr2 along with Fgfr1 which were shown as critical FGF receptors in establishing the primitive endoderm lineage [27]. Thus understanding the molecular determinants that establish this FGF4/FGFR2 signalling axis will shed light on the mechanism that establishes cell fate within the ICM.

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 [28-31]. In the enhancers of these genes, a Sox2 binding motif, CTTTG(A/T)(A/T) [32, 33] is found adjacent to an octamer motif, ATGC(A/T)AA(T/A) [34] with a spacer having zero to three base pairs in between the two motifs. A recent study also enlightened the importance of an enhancer where it was illustrated that gene activation is highly correlated with the presence of an optimal motif [35]. Furthermore, crystallography studies have shown that the Sox2 and Oct4 DNA binding domains heterodimerise on this motif [36]. However, unlike *Oct4*, *Sox2* levels show a dynamic pattern in the preimplantation embryo, in particular, zygotic transcription initiates within the inner cells of the morula [13]. Additionally, Sox2 is known to be an activator of *Fgf4* [37] and a repressor of *Fgfr2* [38]. Importantly, Sox2 is required for normal development as Sox2 null embryos fail to develop beyond early post-implantation [39] and is required non-cell-autonomously via FGF4 for the development of the primitive endoderm [40]. Collectively, these observations indicate that understanding Sox2 dynamics quantitatively is paramount to understanding the molecular mechanism of cell fate decision within the ICM.

We had previously proposed a model based on the dynamics of *Sox2*, *Fgf4*, and *Fgfr2* expression whereby the initiation of Sox2 expression in inner cells of the morula establishes the FGF signalling axis, via the up-regulation of *Fgf4* and the down-regulation of *Fgfr2*, within the ICM [13]. Here we define the *cis* regulatory logic for this model by measuring the dynamic changes in Sox2 levels through preimplantation development and determining the apparent

dissociation constants (aK_d) of Sox2 and Oct4 on their respective *cis* regulatory motifs on target genes of interest. We perform these measurements through the use of fluorescent fusion proteins and fluorescent correlation spectroscopy, a single molecule sensitive fluorescence-based technique [41, 42]. 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 motif does not show such a high dependency on the level of Sox2 compared to that of the *Fgf4* Sox/Oct motif. These biochemical measurements lend weight to the argument that Sox2 is indeed the driver of the earliest heterogeneity within the ICM, a heterogeneity that leads to the EPI/PrE cell fate decision.

MATERIALS AND METHODS

Electrophoretic mobility shift assay (EMSA)- EMSA was carried out as described [31, 43]. Details of quantitative titration assay are provided in **supplementary information**. Additionally a fluorescence protein based EMSA (FP-EMSA) was also performed as described [43, 44]. All the oligonucleotides used for EMSA are provided in the **supplementary information**.

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°C with 5% CO_2. For a 24-well plate, 0.5 µg of EO plasmid, 0.5 µg of MS plasmid and 0.3 µg of pGL3 NSO plasmid were co-transfected per well using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. 0.05 µg of *Renilla* luciferase plasmid (pRL-TK from Promega) was co-transfected as an internal control. Firefly and *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 [31].

Concentration measurement of fusion protein - The concentration of fusion proteins in unpurified nuclear lysate by fluorescence correlation spectroscopy (FCS) was measured as

described [45, 46]. See the **supplementary information** for the theoretical models used in FCS analysis.

Data analysis (EMSA& FCS) for aK_d determination - The apparent dissociation constants, **aK_d**, were determined as described earlier [46].

Immunocytochemical Staining and Image J based semi-quantification - Embryos were fixed in 2.5% PFA for 15 min at 37ºC, washed with Triton (0.1% in PBS; 5 min), Triton (0.5% in PBS; 20min), Triton (0.1% in PBS; 5 min), and BSA/Tween (0.1%BSA and 0.01% Tween in PBS; 30 min). After incubation with 1º antibody (Sox2-Y17) in BSA/Tween (60 min), embryos were washed with BSA/Tween (3x15min), and then incubated with 2º antibody (Goat anti-Rabbit IgG) conjugated with Alexa Fluor 488 (Molecular Probes, Carlsbad, CA) for an additional hour. Following BSA/Tween washes (3x15min), 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 lapse imaging for detection of GFP signal in growing embryos in culture - Two cell-stage embryos, derived by crossing of Sox2-GFP heterozygotic females [47] and males, were cultured in M2 media on the Zeiss microscope (Axio Observer D1, Zeis) stage in appropriate culture conditions (CO2-5%, 37ºC). 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 [47] 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 analysed 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.

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 [13]. To assess the onset of zygotic Sox2 expression, we measured paternally derived zygotic GFP expression from the *Sox2* locus in *Sox2*-null embryos [47] following progressive cell stages during preimplantation development (**Figure 1A**). The earliest expression of GFP was found to be within the inner cells of the morula and later restricted to the ICM of the blastocyst. Next we measured total Sox2 levels by immunostaining (**Figure 1B)**. While Sox2 levels are relatively high at the 4-cell stage levels decrease as development progresses to the morula. Within the blastocyst, Sox2 levels continue to recede in the TE whereas there is an increase in Sox2 from the morula to ICM (**Figure 1C**). This Sox2 protein dynamics closely parallels the dynamics of its mRNA level [13]. The impact of this increasing concentration of Sox2 from the morula-ICM transition will be mediated through *cis* regulatory logic**.**

Characterization of the Sox/Oct motif sequences – Despite the increasing levels of Sox2 in the nascent ICM some known target genes of Sox2 did not change (e.g. *Nanog)* while others did (e.g. *Fgf4*) [13], thus we were next interested in determining if this could be explained through specific variations in *cis* regulatory sequences mediating Sox2 binding. From a global view of Sox2 and Oct4 bound regions in embryonic stem (ES) cells it is clear that while there is sequence constraint within the Sox/Oct motif, there is allowable variability (**Figure 2A**). This variability is in contrast to the sequence conservation seen within the particular Sox/Oct motif in both *Nanog* and *Fgf4*, where sequence conservation verges on 100% identity over 100s of millions of years of cumulative evolution (**Figure 2B**). Furthermore, we also observed the

conservation pattern of different Sox/Oct motifs from the known genes, *Nanog, Utf1, Oct4, Sox2, and Fgf4* (**Figure 2C**). Such sequence conservation strongly argues that there are functional differences between sequences that encompass the allowable Sox/Oct motif.

Through site-directed mutagenesis of the *Nanog* Sox/Oct motif, within the larger context of the *Nanog* 400 bp proximal promoter, we tested the functional consequences of subtle mutations on transcription as measured by luciferase activity generated in transfected F9 teratocarcinoma cells. As previously described [31], ablation of the binding sites for Sox2 or Oct4, and in combination, through 3 bp mutations diminishes luciferase activity to below 20% of the wild-type promoter (**Figure 2D,E**). When we introduced subtle single base changes of A to T, T to A, and A to T at positions 2, 6, and 7 of the sox motif, respectively, there was a significant reduction in luciferase activity in all cases from 30 to 60% of wild-type levels (**Figure 2D,E**). Particularly surprising was the reduction to ~60% at position 7 of the sox motif as this is the least conserved of all seven positions within the sox motif (**Figure 2A,C**). Thus what apparently are subtle changes to the Sox/Oct motif that qualitatively do not prevent binding of Sox2, have profound functional consequences on transcriptional output. We would argue that such functional consequences result in the high level of sequence conservation, through purifying selection, within specific Sox/Oct motifs across eutherian species, particularly around these developmental control genes.

Characterization of TFs-fluorescent fusion proteins – We hypothesized that the differential transcriptional response to increasing concentrations of Sox2 in the nascent ICM seen between *Nanog* and *Fgf4* was a result of differential binding kinetics of Sox2 and Oct4 on the associated *Sox/Oct* motifs of these genes. As no technologies exist to measure protein-DNA binding kinetics at discrete genomic loci within living cells, we resorted to *in vitro* measurements. Our strategy was to generate quantitative measurements through Förster resonance energy transfer (FRET), fluorescence correlation spectroscopy (FCS) and electrophoretic mobility shift assay (EMSA) on Sox2-Oct4-DNA complexes and thus required the generation of fluorescently tagged proteins.

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 transcription factor-fluorescent protein fusion constructs, and with nuclear localization (**Figure 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 (**Figure 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 (**Supplementary Figure S1**). 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 (**Figure 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, by EMSA, nuclear lysates from ES cells transfected with these plasmid constructs for their ability to bind an oligonucleotide containing the *Nanog* Sox/Oct motif. We observed that both endogenous proteins and their fusion counterparts are capable of binding this DNA motif as monomers and heterodimers (**Figure 3D**). Differentially shifted bands indicate 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 (**Figure 3B**) and requiring to compete with their endogenous counterparts in this assay. These results indicate these fusion proteins are as competent as their wild type proteins in binding to DNA containing the Sox/Oct motif. Finally, we confirmed the physiological function of the N-terminal fusion proteins, GFP-Oct4 and mCherry-Sox2 for their ability to rescue ES cells in which the corresponding endogenous TF alleles had been deleted [46].

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. Utilizing 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. No FRET signal was observed from a solution containing GFP-Oct4 and mCherry-Sox2 in the absence of DNA (**Figure 3E**) however, when DNA containing a *Nanog* Sox/Oct motif was included a distinct FRET signal was detected (**Figure 3F**). This observation indicates that the DNA brings GFP-Oct4 and mCherry-Sox2 into close proximity enabling successful energy transfer from GFP to mCherry, through binding to the Sox/Oct motif. In further validation, our FP-EMSA assay also did not detect Sox2-Oct4 interaction unless Sox/Oct DNA was present **(Figure 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 [43, 46].

Determination of apparent dissociation constants (aK_d) 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 (**Supplementary Table 1)**. For these quantitative studies we used FCS and EMSA as complementary methods. FCS has been used for a number of years to measure aK_ds in lysate, live cells, and zebrafish embryos [46, 48]. More recently we have established FP-EMSA as a complementary technique for quantitative aK_d measurements using full-length fusion proteins [46]. The titration strategy is shown in **Figure 4A**. Direct evidence of complex formation on the *Fgf4* Sox/Oct motif under different titration conditions was compared between EMSA generated gel images and FCS generated ACF curves (**Figure 4B**). In the presence of 72 nM mCherry-Sox2, a titration of GFP-Oct4 to the *Fgf4* Sox/Oct motif yielded aK_ds of 25.2 ± 4.1 nM and 25.3 ± 2.2 nM from EMSA and FCS, respectively (**Figure 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 aK_d of 23.2 ± 1.2 nM and aK_d of 24.0 ± 3.0 nM from EMSA and FCS, respectively (**Figure 4D**). Notably, both FCS and EMSA provided similar values within the margins of standard deviation, strengthening the reliability of our quantitative findings on TF-DNA binding interactions.

Influence of sequence variation within the Sox/Oct motif on protein-DNA binding affinity - Previously we had observed that sequence changes in the conserved and non-conserved regions of the Sox2 motif had an impact on transcriptional activity (**Fig. 2E**). Such activity may be linked to the degree of Oct4 and Sox2 binding affinities for these DNA motifs

(**Supplementary Table 1)** and consequently impact the formation of a stable heterodimer. The change in the sequence of the Sox2 binding site "CATTGTA" in *Nanog* to "CATTGTT" 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 7th position of the sequence "CATTGTA" 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 2nd and 7th positions were changed to "CTTTGTT" in *Fgf4* from "CATTGTA" in *Nanog* corresponding to an increase in aK_d to ~70 nM (**Table 1 & Supplementary Figure S2**). Our result demonstrates that the variable positions in the heptamer sequence play an important role in the binding interactions of Sox2 with the Sox/Oct motif while the conserved positions are 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 (**Figure 2C**) we know that this conserved region has an important role in Oct4 binding to its motif; but the non-conserved region (5th to 8th 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 7th 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 7th position plays an important role in increasing the aK_d value from 32.0 ± 5.5 nM to 42.5 ± 5.5 nM. The 5th and 6th 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 (**Table 1 & Supplementary Figure S2**). 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 and thus could be anticipated that ATGC has more influence on Oct4 binding specificity to the octamer sequence of Sox/Oct motifs. Therefore, in addition to 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) than the last 4 base pairs (AAAA) (**Supplementary Figure S3A & Supplementary Table 2**). We further attempted 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 (**Supplementary Figure S3B & Supplementary Table 2**). These results indicate the Sox/Oct motif with a base change 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 concentration on 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 (**Figure 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, there was a significant difference in stable ternary complex recruitment among the *Fgf4*, *Nanog* and *Utf1* Sox/Oct motifs whereby the *Fgf4* Sox/Oct motif required higher concentrations 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 the *Nanog* motif 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, the *Nanog* and *Utf1* elements showed a similar 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 the *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 preimplantation development [13]. Consistent with this observation, we too found that the protein levels of Sox2 showed similar trend as seen earlier (**Figure 1**).

Comparison of Sox2 binding affinity between *Fgf4* and *Fgfr2 cis* regulatory motifs – Sox2 is bound to an intronic region of *Fgfr2* in mouse embryonic stem cells [49]. We analysed this data and identified a canonical Sox *cis* motif within this ChIP-seq peak (**Figure 5A and Supplementary figure S4**). We analysed the extent of sequence conservation within this Sox *cis* motif and found it identical in mouse and rats but altered beyond a recognizable Sox motif in human and chimp. This suggests that this is a functional Sox2 binding motif in murine but not in the human genome. Masui et al. (2007) had experimentally shown that the removal of Sox2 in mouse ES cells results in *Fgfr2* upregulation, thus suggesting Sox2 may be repressing *Fgfr2* expression through this intronic Sox element [38]. We next sought to determine whether Sox2 interacts with this motif in a similar fashion as the *Fgf4* Sox/Oct motif as, in our proposed, model both *Fgf4* and *Fgfr2* expression are sensitive to varying Sox2 concentrations [13]. We therefore 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 motifs (**Figure 5B**). We further quantified the binding affinities of Sox2 to these motifs by FCS (**Figure 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 (**Figure 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 1**). This could be the reason for the good correlation between *Fgf4* expression and Sox2 levels 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 *in vivo*. We next sought to answer this through analysis of Sox2-null embryos.

Validation of the role of Sox2 on its target genes in Sox2-null embryos by RT-qPCR - Taken together, our present and published work [13] suggests that Sox2 works as regulator for both *Fgf4* (positively) and *Fgfr2* (negatively). Zygotic Sox2 expression regulates early

embryonic development [39]. To assess the expression of genes confirmed to possess Sox2 interacting regulatory elements, quantitative mRNA analysis was performed in Sox2-null and wild-type embryos, obtained from intercrosses of $Sox2^{GFP}$ heterozygotes [47], after first normalizing expression levels to β-actin. This showed that the Epi marker *Fgf4* was down-regulated significantly in Sox2-null embryos and the PrE marker *Fgfr2* was up-regulated slightly (**Figure 6B**). We also observed that expression of *Nanog* is not influenced by the absence of Sox2 which further supports our *in vitro* data where we observed that the *Nanog* Sox/Oct motif is capable of making stable ternary complexes at even lower Sox2 concentrations than the *Fgf4* Sox/Oct motif. It should be noted that maternal Sox2 is still present in the early blastocyst [39] and this could be sufficient for the expression of *Nanog* but not for that of *Fgf4* and *Fgfr2* as they require higher Sox2 levels.

DISCUSSION

Whether the segregation of ICM into PE and EPI is driven by stochastic or deterministic events is currently controversial [9, 50] . However, this differentiation event depends on Fgf/Erk signalling which in turn depends upon communication between inner cells expressing *Fgf4* and inner cells expressing *Fgfr2* [9, 13, 22, 23]. 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 zygotic 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 (**Figure 6A**). At 37 nM Sox2, the aK_d 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 aK_d 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* expression may be responsive to Sox2 while *Fgf4* expression may not. This agrees with mRNA expression analysis in Sox2 null embryos (**Figure 6B**) at E3.5, showing that *Nanog* is expressed but *Fgf4* is not. At 72 nM Sox2, the aK_d 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 (**Figure 6C**) controlling the segregation of the ICM into the EPI and the PE. As zygotic Sox2 expression is first detectable in the 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 (derived from earlier outer cells) 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 [51].

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.

Overall, our working model supports the previously reported 'time-inside time-outside' [50-52] or 'integrated' cell-fate decision model [53] in which ICM cells generated as a result of the 8- to 16-cell transition are biased to form EPI whereas those generated during the morula to blastocyst (16- to 32-cell) transition are biased to form PE. Our results are consistent with previously reported relative expression levels of *Fgf4* and *Fgfr2* transcripts in inner cells derived from the 8- to 16- and 16- to 32-cell stage transitions [26] and with a non-cell autonomous role of Sox2 promoting PE development via Fgf4 [40]. In contrast, our findings are not consistent with reports that suggest cell history has little influence on the cell fate decision segregating PE and EPI [18, 22]. Notably however, while Ohnishi et al (2014) argue this cell fate decision is stochastic, they do find bimodal expression of *Fgf4* within the earliest ICM stage they analysed [22]. We argue this bimodal expression of *Fgf4* is a result of inner cell-specific initiation of zygotic Sox2 expression combined with the sensitivity of the Sox/Oct *cis* regulatory motif within *Fgf4* to Sox2 levels. This second cell fate decision is linked to the first cell fate decision via hippo signalling as expression of zygotic Sox2 in inner cells requires active hippo signalling, which in turn is controlled by the relative positioning of cells within the morula [40, 54-58]. It is also worth mentioning the Sox2-bound, repressively acting *cis* motif we identify in *Fgfr2* is conserved in rodents, but unrecognizable in humans - this provides a feasible explanation for the observation that cell fate decision within the human ICM is independent of Fgf signalling [59].

In summary, the *in vitro* protein-DNA binding data and *in vivo* analysis of Sox2 levels controlling the expression of *Fgf4* and *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.

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 and Ms. Indrani Mistri Jana for proof reading.

AUTHOR CONTRIBUTIONS

T.K.M did experiments, 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, L.H.L. did the promoter mutagenesis and luciferase assays. I.C. supervised aspects of the work, which overall was supervised by T.W. and P.R. T.K.M, I.C., T.W. and P.R. wrote the manuscript.

FUNDING

Work in IC's lab was funded by the Medical Research Council of the United Kingdom and by the Human Frontier Sciences Program. TKM was the recipient of an NUS research scholarship. TW and PR gratefully acknowledge funding by a National University of Singapore/Baden-Württemberg (NUS-BW), R-143-000-422-646 joint grant. Additional research in the PR lab was supported by the Agency for Science Technology & Research (A*STAR).

COMPETING INTERESTS

The authors declare that there are no competing interests associated with the manuscript.

ADDITIONAL INFORMATION

Please see the **supplementary information** for further information.

REFERENCE

1 Rossant, J. and Tam, P. P. L. (2009) Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. Development. **136**, 701-713

2 Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S. and Smith, A. (2003) Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell. **113**, 643-655

3 Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M. and Yamanaka, S. (2003) The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell. **113**, 631-642

4 Silva, J., Nichols, J., Theunissen, T. W., Guo, G., van Oosten, A. L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I. and Smith, A. (2009) Nanog is the gateway to the pluripotent ground state. Cell. **138**, 722-737

5 Morrisey, E. E., Ip, H. S., Lu, M. M. and Parmacek, M. S. (1996) GATA-6: a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. Developmental biology. **177**, 309-322

6 Morrisey, E. E., Tang, Z., Sigrist, K., Lu, M. M., Jiang, F., Ip, H. S. and Parmacek, M. S. (1998) GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. Genes & development. **12**, 3579-3590

7 Koutsourakis, M., Langeveld, A., Patient, R., Beddington, R. and Grosveld, F. (1999) The transcription factor GATA6 is essential for early extraembryonic development. Development. **126**, 723-732

8 Aksoy, I., Jauch, R., Chen, J., Dyla, M., Divakar, U., Bogu, G. K., Teo, R., Ng, C. K. L., Herath, W. and Lili, S. (2013) Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. The EMBO journal. **32**, 938-953

9 Chazaud, C., Yamanaka, Y., Pawson, T. and Rossant, J. (2006) Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. Developmental cell. **10**, 615-624

10 Xenopoulos, P., Kang, M., Puliafito, A., Di Talia, S. and Hadjantonakis, A.-K. (2015) Heterogeneities in Nanog expression drive stable commitment to pluripotency in the mouse blastocyst. Cell reports. **10**, 1508-1520

11 Schrode, N., Saiz, N. s., Di Talia, S. and Hadjantonakis, A.-K. (2014) GATA6 levels modulate primitive endoderm cell fate choice and timing in the mouse blastocyst. Developmental cell. **29**, 454-467

12 Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H. and Smith, A. (1998) Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell. **95**, 379-391

13 Guo, G., Huss, M., Tong, G. Q., Wang, C., Li Sun, L., Clarke, N. D. and Robson, P. (2010) Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Developmental cell. **18**, 675-685

14 Krawchuk, D., Honma-Yamanaka, N., Anani, S. and Yamanaka, Y. (2013) FGF4 is a limiting factor controlling the proportions of primitive endoderm and epiblast in the ICM of the mouse blastocyst. Developmental biology. **384**, 65-71

15 Niswander, L. and Martin, G. R. (1992) Fgf-4 expression during gastrulation, myogenesis, limb and tooth development in the mouse. Development. **114**, 755-768

16 Orr-Urtreger, A., Givol, D., Yayon, A., Yarden, Y. and Lonai, P. (1991) Developmental expression of two murine fibroblast growth factor receptors, flg and bek. Development. **113**, 1419-1434

17 Nichols, J., Silva, J., Roode, M. and Smith, A. (2009) Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. Development. **136**, 3215-3222

18 Yamanaka, Y., Lanner, F. and Rossant, J. (2010) FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. Development. **137**, 715
19 Thamodaran, V. and Bruce, A. W. (2016) p38 (Mapk14/11) occupies a regulatory node governing entry into primitive endoderm differentiation during preimplantation mouse embryo development. Open biology. **6**, 160190
20 Feldman, B., Poueymirou, W., Papaioannou, V. E., DeChiara, T. M. and Goldfarb, M. (1995) Requirement of FGF-4 for postimplantation mouse development. Science. **267**, 246
21 Arman, E., Haffner-Krausz, R., Chen, Y., Heath, J. K. and Lonai, P. (1998) Targeted disruption of fibroblast growth factor (FGF) receptor 2 suggests a role for FGF signaling in pregastrulation mammalian development. Proceedings of the National Academy of Sciences. **95**, 5082
22 Ohnishi, Y., Huber, W., Tsumura, A., Kang, M., Xenopoulos, P., Kurimoto, K., OleÅ›, A. K., AraÃºzo-Bravo, M. J., Saitou, M. and Hadjantonakis, A.-K. (2014) Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. Nature cell biology. **16**, 27-37
23 Kang, M., Piliszek, A., Artus, J. r. m. and Hadjantonakis, A.-K. (2013) FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their initial expression in the mouse. Development. **140**, 267-279
24 Fleming, T. P. (1987) A quantitative analysis of cell allocation to trophectoderm and inner cell mass in the mouse blastocyst. Developmental biology. **119**, 520-531
25 Pedersen, R. A. (1986) Potency, lineage, and allocation in preimplantation mouse embryos. Cambridge University Press, Cambridge
26 Krupa, M., Mazur, E., Szczepańska, K., Filimonow, K., Maleszewski, M. and Suwińska, A. (2014) Allocation of inner cells to epiblast vs primitive endoderm in the mouse embryo is biased but not determined by the round of asymmetric divisions (8⟶ 16-and 16⟶ 32-cells). Developmental biology. **385**, 136-148
27 Kang, M., Garg, V. and Hadjantonakis, A.-K. (2017) Lineage establishment and progression within the inner cell mass of the mouse blastocyst requires FGFR1 and FGFR2. Developmental Cell
28 Nishimoto, M., Fukushima, A., Okuda, A. and Muramatsu, M. (1999) The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. Molecular and cellular biology. **19**, 5453
29 Ambrosetti, D. C., Schöler, H. R., Dailey, L. and Basilico, C. (2000) Modulation of the Activity of Multiple Transcriptional Activation Domains by the DNA Binding Domains Mediates the Synergistic Action of Sox2 and Oct-3 on the Fibroblast Growth Factor-4Enhancer. Journal of Biological Chemistry. **275**, 23387
30 Chew, J. L., Loh, Y. H., Zhang, W., Chen, X., Tam, W. L., Yeap, L. S., Li, P., Ang, Y. S., Lim, B. and Robson, P. (2005) Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. Molecular and cellular biology. **25**, 6031
31 Rodda, D. J., Chew, J. L., Lim, L. H., Loh, Y. H., Wang, B., Ng, H. H. and Robson, P. (2005) Transcriptional regulation of nanog by OCT4 and SOX2. Journal of Biological Chemistry. **280**, 24731
32 Harley, V. R., Lovell-Badge, R. and Goodfellow, P. N. (1994) Definition of a consensus DNA binding site for SRY. Nucleic acids research. **22**, 1500
33 Wilson, M. and Koopman, P. (2002) Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. Current opinion in genetics & development. **12**, 441-446
34 Verrijzer, C. P., Alkema, M. J., van Weperen, W. W., Van Leeuwen, H. C., Strating, M. J. and Van der Vliet, P. C. (1992) The DNA binding specificity of the bipartite POU domain and its subdomains. The EMBO Journal. **11**, 4993
35 Farley, E. K., Olson, K. M., Zhang, W., Brandt, A. J., Rokhsar, D. S. and Levine, M. S. (2015) Suboptimization of developmental enhancers. Science. **350**, 325-328
36 Reményi, A., Lins, K., Nissen, L. J., Reinbold, R., Schöler, H. R. and Wilmanns, M. (2003) Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. Genes & development. **17**, 2048

37 Yuan, H., Corbi, N., Basilico, C. and Dailey, L. (1995) Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. Genes & development. **9**, 2635-2645
38 Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R. and Sharov, A. A. (2007) Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nature cell biology. **9**, 625-635
39 Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N. and Lovell-Badge, R. (2003) Multipotent cell lineages in early mouse development depend on SOX2 function. Genes & development. **17**, 126
40 Wicklow, E., Blij, S., Frum, T., Hirate, Y., Lang, R. A., Sasaki, H. and Ralston, A. (2014) HIPPO pathway members restrict SOX2 to the inner cell mass where it promotes ICM fates in the mouse blastocyst.
41 Elson, E. L. and Magde, D. (1974) Fluorescence correlation spectroscopy. I. Conceptual basis and theory. Biopolymers. **13**, 1-27
42 Jameson, D. M., Ross, J. A. and Albanesi, J. P. (2009) Fluorescence fluctuation spectroscopy: ushering in a new age of enlightenment for cellular dynamics. Biophysical Reviews. **1**, 105-118
43 Chen, S. L., Tapan, K. M., Yong, H. F., Thankiah, S., Hui, T. G., David, R., Leng, H. L., Chai, C., Paul, R. and Thorsten, W. (2012) DNA-dependent Oct4-Sox2 interaction and diffusion properties characteristic of the pluripotent cell state revealed by fluorescence spectroscopy. Biochemical Journal. **448**, 21-33
44 Hutchins, A. P., Choo, S. H., Mistri, T. K., Rahmani, M., Woon, C. T., Keow Leng Ng, C., Jauch, R. and Robson, P. (2013) Co-motif discovery identifies an Esrrb-Sox2-DNA ternary complex as a mediator of transcriptional differences between mouse embryonic and epiblast stem cells. Stem Cells. **31**, 269-281
45 Buschmann, V., Krämer, B., Koberling, F., Macdonald, R. and Rüttinger, S. (2009) Quantitative FCS: Determination of the Confocal Volume by FCS and Bead Scanning with the MicroTime 200. PicoQuant GMbH
46 Mistri, T. K., Devasia, A. G., Chu, L. T., Ng, W. P., Halbritter, F., Colby, D., Martynoga, B., Tomlinson, S. R., Chambers, I. and Robson, P. (2015) Selective influence of Sox2 on POU transcription factor binding in embryonic and neural stem cells. EMBO reports. **16**, 1177-1191
47 Ellis, P., Fagan, B. M., Magness, S. T., Hutton, S., Taranova, O., Hayashi, S., McMahon, A., Rao, M. and Pevny, L. (2004) SOX2, a persistent marker for multipotential neural stem cells derived from embryonic stem cells, the embryo or the adult. Developmental neuroscience. **26**, 148-165
48 Shi, X., Foo, Y. H., Sudhaharan, T., Chong, S.-W., Korzh, V., Ahmed, S. and Wohland, T. (2009) Determination of dissociation constants in living zebrafish embryos with single wavelength fluorescence cross-correlation spectroscopy. Biophysical journal. **97**, 678-686
49 Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W. and Jiang, J. (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell. **133**, 1106-1117
50 Morris, S. A., Teo, R. T. Y., Li, H., Robson, P., Glover, D. M. and Zernicka-Goetz, M. (2010) Origin and formation of the first two distinct cell types of the inner cell mass in the mouse embryo. Proceedings of the National Academy of Sciences. **107**, 6364-6369
51 Morris, S. A., Graham, S. J., Jedrusik, A. and Zernicka-Goetz, M. (2013) The differential response to Fgf signalling in cells internalized at different times influences lineage segregation in preimplantation mouse embryos. Open biology. **3**, 130104
52 Morris, S. A. (2011) Cell fate in the early mouse embryo: sorting out the influence of developmental history on lineage choice. Reproductive biomedicine online. **22**, 521-524
53 Mihajlović, A. I., Thamodaran, V. and Bruce, A. W. (2015) The first two cell-fate decisions of preimplantation mouse embryo development are not functionally independent. Scientific reports. **5**, 15034

54 Hirate, Y., Hirahara, S., Inoue, K.-i., Suzuki, A., Alarcon, V. B., Akimoto, K., Hirai, T., Hara, T., Adachi, M. and Chida, K. (2013) Polarity-dependent distribution of angiomotin localizes Hippo signaling in preimplantation embryos. Current Biology. **23**, 1181-1194

55 Leung, C. Y. and Zernicka-Goetz, M. (2013) Angiomotin prevents pluripotent lineage differentiation in mouse embryos via Hippo pathway-dependent and-independent mechanisms. Nature communications. **4**, 2251

56 Kono, K., Tamashiro, D. A. A. and Alarcon, V. B. (2014) Inhibition of RHO–ROCK signaling enhances ICM and suppresses TE characteristics through activation of Hippo signaling in the mouse blastocyst. Developmental biology. **394**, 142-155

57 Mihajlović, A. I. and Bruce, A. W. (2016) Rho-associated protein kinase regulates subcellular localisation of Angiomotin and Hippo-signalling during preimplantation mouse embryo development. Reproductive biomedicine online. **33**, 381-390

58 Sasaki, K., Kakuwa, T., Akimoto, K., Koga, H. and Ohno, S. (2015) Regulation of epithelial cell polarity by PAR-3 depends on Girdin transcription and Girdin–Gαi3 signaling. J Cell Sci. **128**, 2244-2258

59 Roode, M., Blair, K., Snell, P., Elder, K., Marchant, S., Smith, A. and Nichols, J. (2012) Human hypoblast formation is not dependent on FGF signalling. Developmental biology. **361**, 358-363

60 Sharov, A. A. and Ko, M. S. H. (2009) Exhaustive search for over-represented DNA sequence motifs with CisFinder. DNA research. **16**, 261-273

61 Chambers, I. (2005) Mechanisms and factors in embryonic stem cell self-renewal. Rendiconti Lincei. **16**, 83-97

62 Longabaugh, W. J. (2012) BioTapestry: A Tool to Visualize the Dynamic Properties of Gene Regulatory Networks. Methods in molecular biology (Clifton, NJ). **786**, 359

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 [13] 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. For each panel, left: GFP, middle: bright field, right: merge. With the progress of experiment, embryos generally move and therefore, panel size was kept bigger in later stage of the experiment keeping the scale bar same. Scale bars are 20μm in size in all panels. No of Samples (S) = 5, No of Replicates (R) = 2.

(**b**) Confocal Z-stack images of representative Sox2 and ToPro nuclear stained embryos. S = 4, R = 2.

(**c**) Quantification of the average Sox2 level per cell in 4-16 cell embryos as well as measurements made at the 64 cell stage were from ICM (red triangle) or TE (yellow circle) cells. S = 4, R = 2, mean ± SEM.

FIGURE 2. **Characterisation of Sox/Oct motifs**.

(a) The top *de novo* sequence motifs (based on enrichment) detected by CisFinder [60] in Oct4 and Sox2 ChIP-Seq data from ESCs [49], shown by WebLogo.

(b) Comparison of *Nanog and Fgf4* Sox/Oct 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) [31, 61] were subsequently tested.

(e) Luciferase activity of constructs shown in (d) after transfection of F9 embryonal carcinoma cells. R = 3, mean ± SD.

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. R = 3.

(**b**) Immunoblot analysis of exogenous fusion proteins; bands A and B correspond to the fusion and wild type proteins respectively. R = 3.

(**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). R = 2, mean ± SD.

(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 motif); B – heterodimer complex (GFP-Oct4/wild type Sox2/NSO motif); C – monomer complex (mCherry-Sox2/NSO motif); D – monomer complex (GFP-Oct4/NSO motif). R = 3.

(**e-f**) FRET experiments to investigate the interaction of Sox2 and Oct4. As donor we use eGFP-Oct4 and as acceptor mCherry-Sox2. Emission intensity is collected at >500 nm wavelength upon GFP excitation by a 488 nm laser line. 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. As indicated in the figure, enhanced emission by mCherry as donor (see arrows at ~610 nm) is only evident in the presence of DNA (f) implying that Oct4 and Sox2 do only interact via DNA. As a positive control we used tandem dye eGFP-mCherry; as negative control we used co-expression of individual eGFP and mCherry proteins. R = 2.

(**g**) Complex formation shown by FP-EMSA detecting GFP-Oct4 in the presence or absence of unlabelled DNA and mCherry-Sox2. R = 3.

FIGURE 4. **Comparison of FCS and EMSA derived aK_d.**

(**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 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. R = 3.

(**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). R = 3, mean ± SEM.

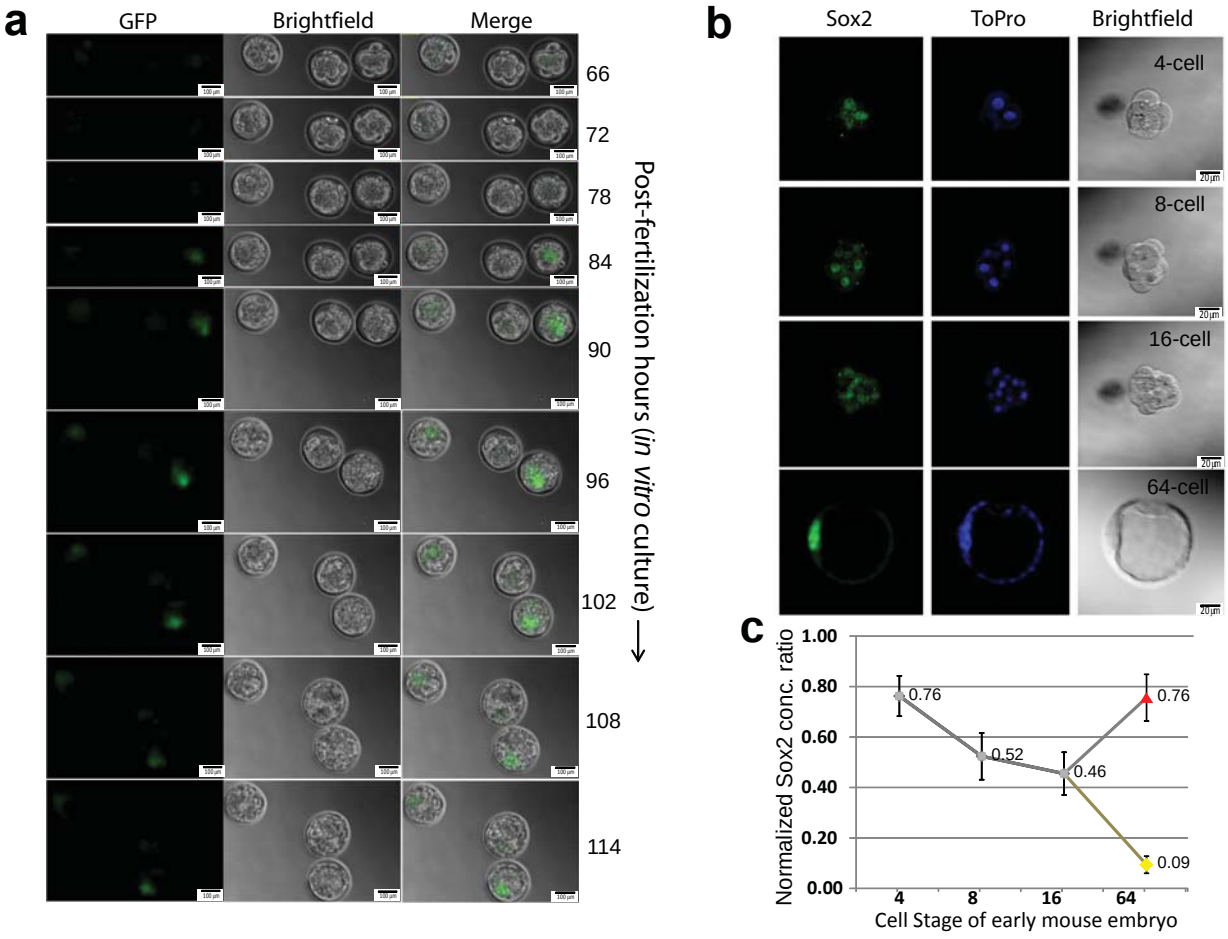
FIGURE 5. **Direct binding of Sox2 protein to *cis* motifs of *Fgf4* and *Fgfr2*.**

(**a**) Identification of the Sox *cis* regulatory motif near the transcription start site of *Fgfr2* from Sox2 ChIP-Seq data in ES cells [49]. 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 motif; lane 1: *Fgfr2* motif, Lane 2: *Fgf4* motif. Yellow and white stars indicate Sox2-DNA complex and free DNA respectively. R = 3
(**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) motifs. The measured aK_ds from each titration assay measured by FCS technique were compared. R = 3, mean ± SEM.

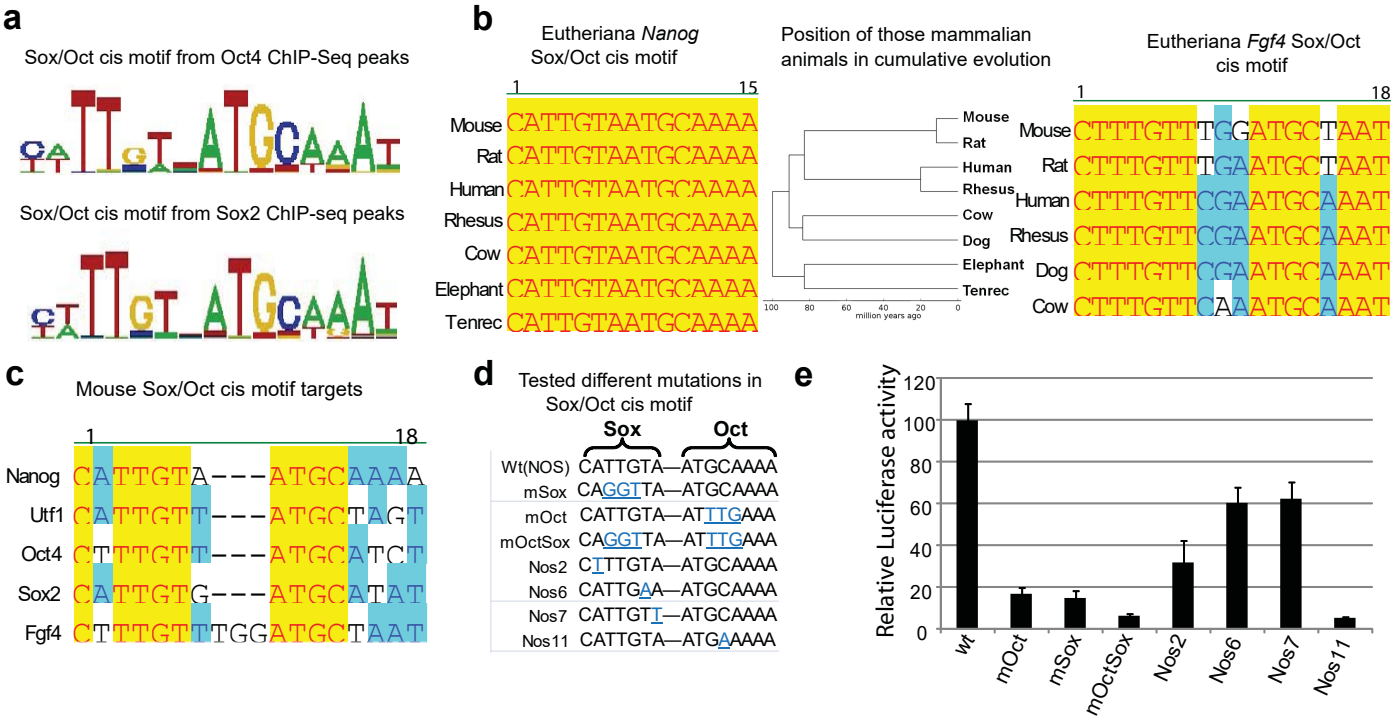
FIGURE 6. **Bridging in *vitro* measurements with *in vivo* Sox2 levels of early embryos.**
(**a**) Hypothetical diagram showing relation of aK_d to gene expression with respect to Sox2 concentration is depicted.
(**b**) Relative mRNA levels of *Fgf4*, *Nanog*, and *Fgfr2* in *Sox2*-null embryos at E3.5 were compared to that of WT embryos by RT-qPCR. S = 4, R = 3, mean ± SD.
(**c**) Gene regulation model for *Fgf4*, *Nanog* and *Fgfr2* during the segregation of ICM into PE and EPI by Bio Tapestry software [62]. 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**. aK_d measurements from different titrations of GFO-Oct4 against *Nanog*, *Fgf4*, and *Utf1* motifs in the presence and absence of mCherry-Sox2 by FCS and vice versa. S = 10, R = 3, mean ± SD.

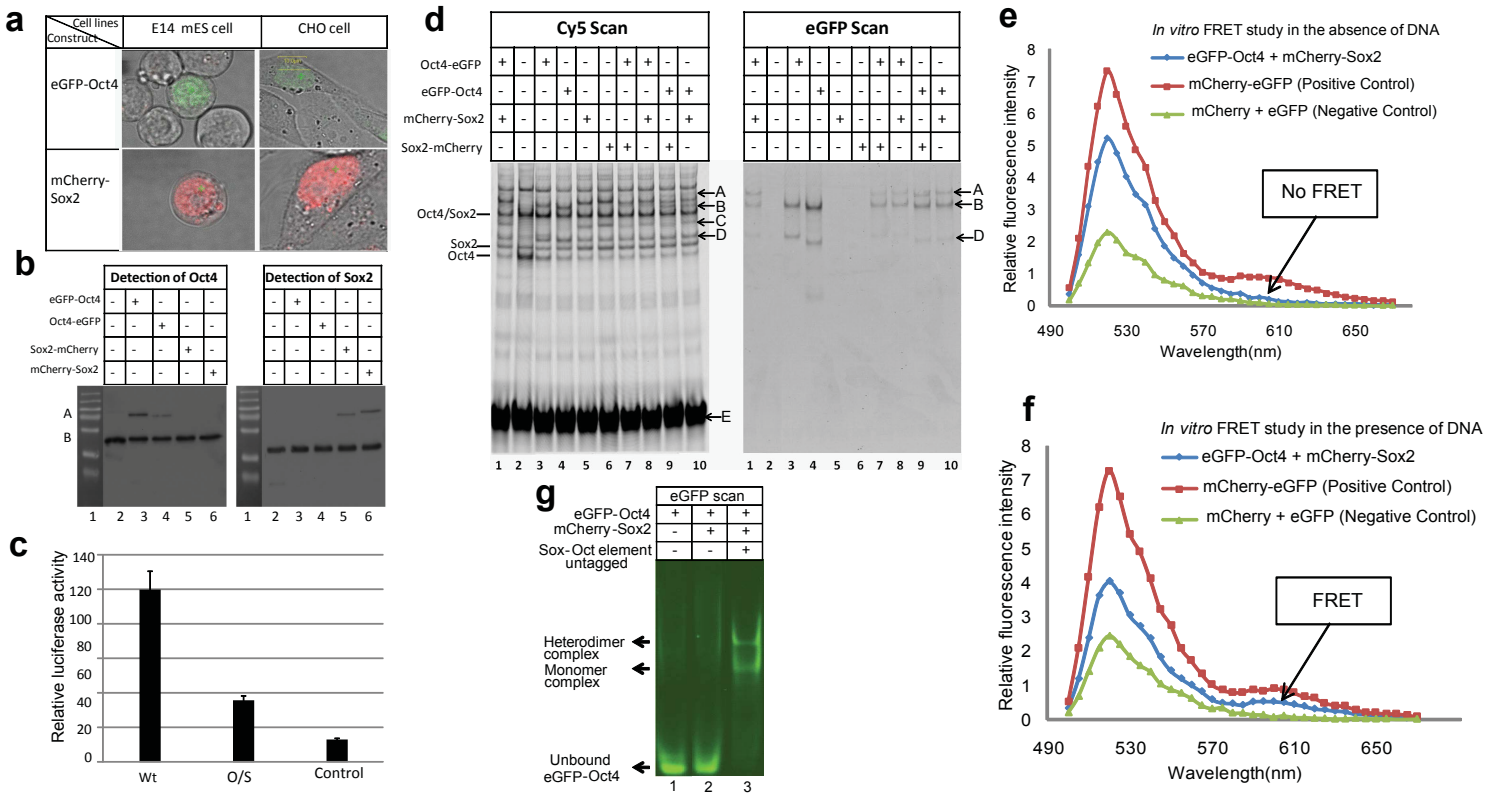
Mistri et al
Figure 1



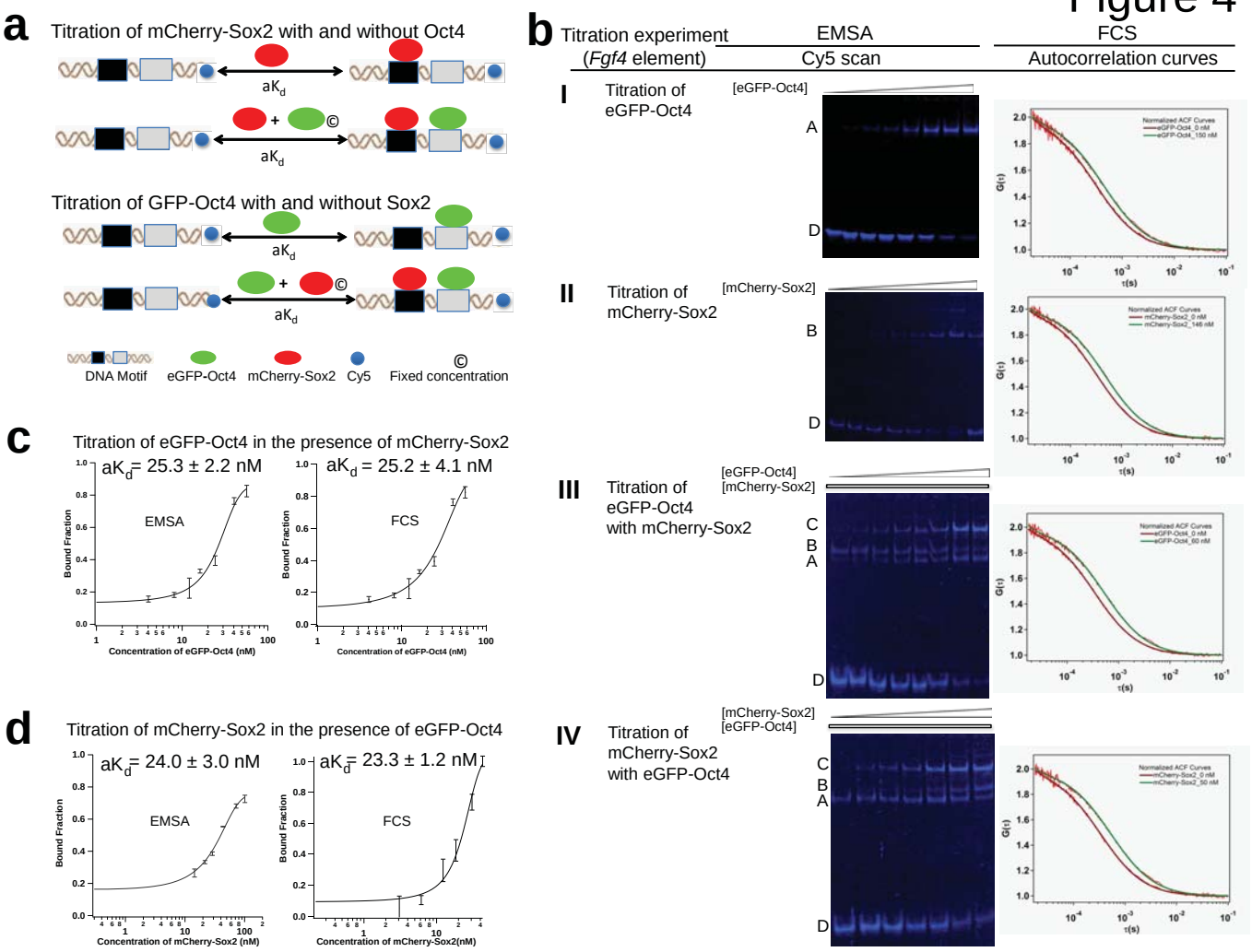
Mistri et al
Figure 2



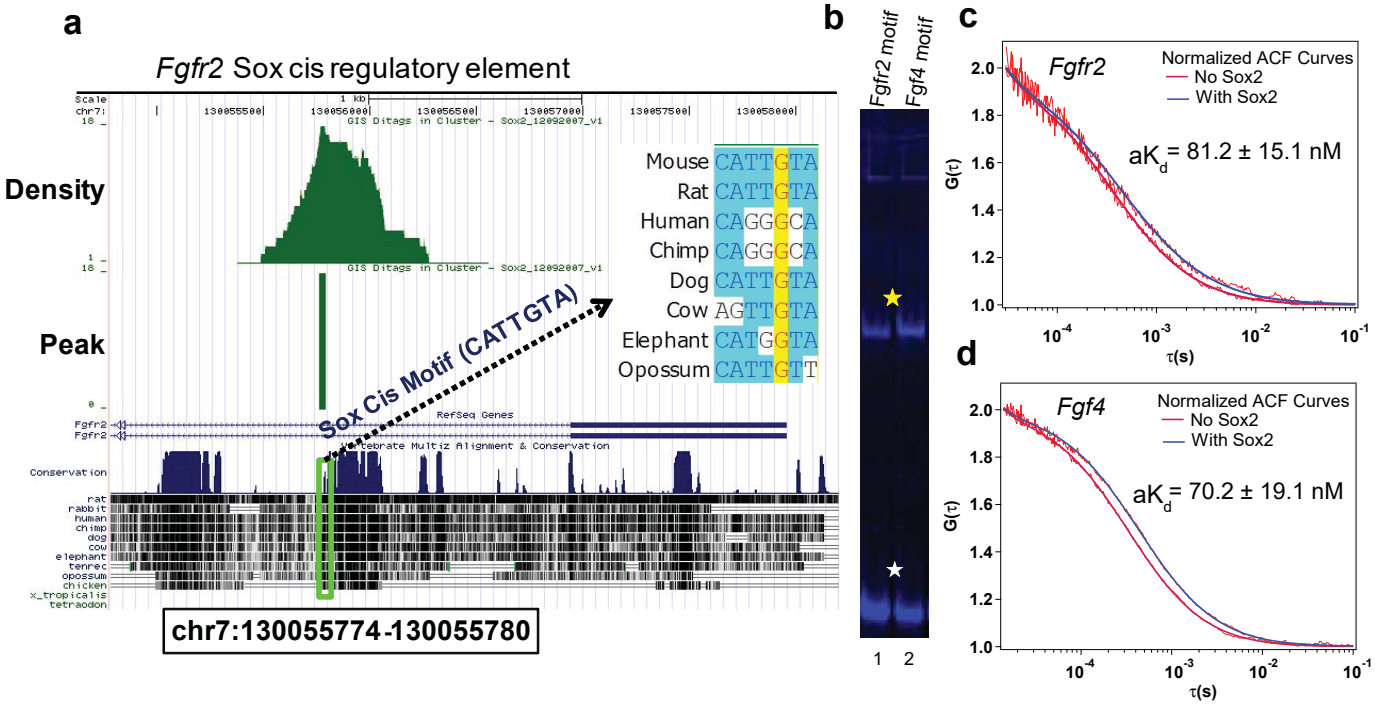
Mistri et al
Figure 3



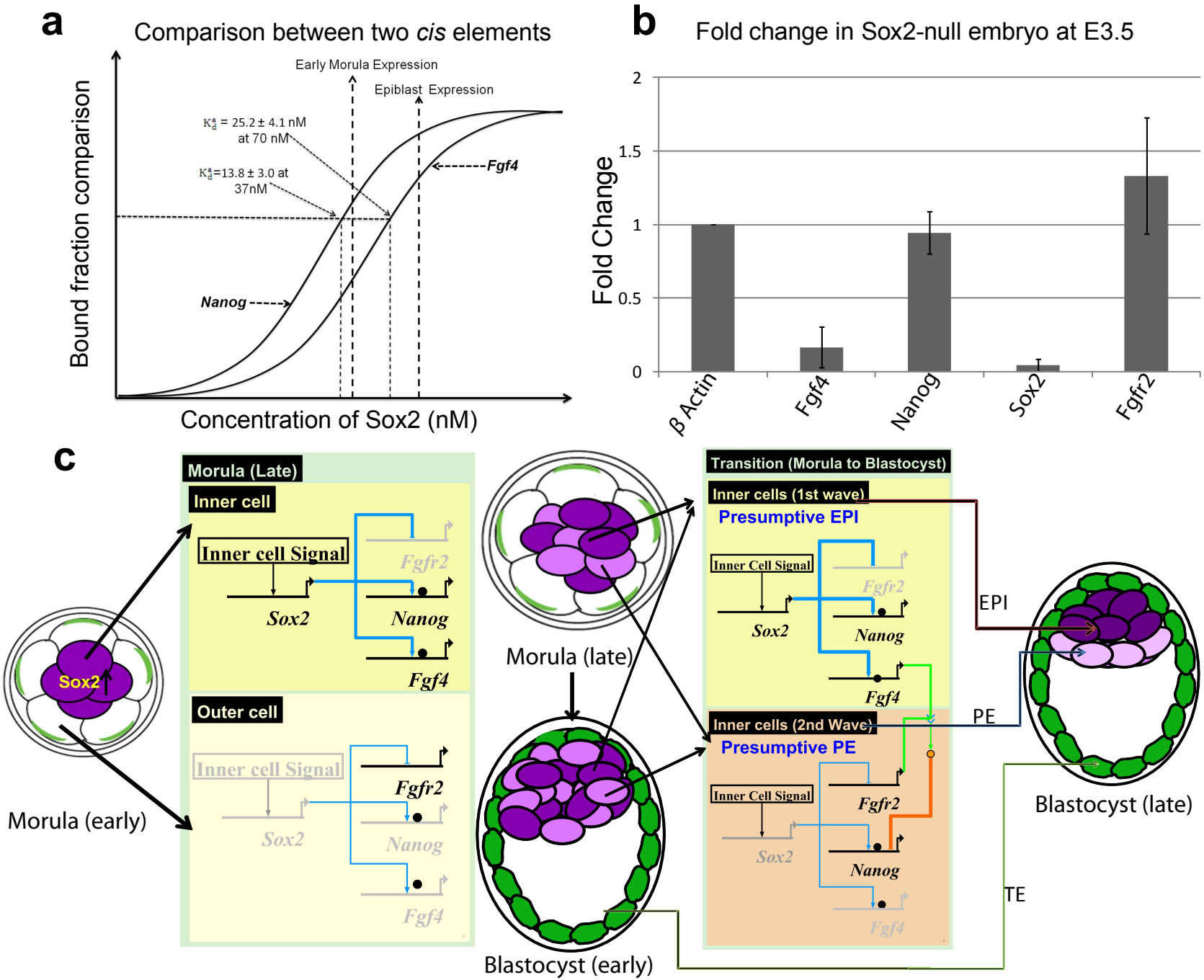
Mistri et al
Figure 4



Mistri et al
Figure 5



Mistri et al
Figure 6



Mistri et al
Table 1

Sox-Oct element	Titration with GFP-Oct4			Titration with mCherry-Sox2	
	Apparent dissociation constant, aK_d (nM)				
	No Sox2	Sox2 at 37nM	Sox2 at 72nM	No Oct4	Oct4 at 40 nM
Nanog element	28.2 ± 4.9	13.8 ± 3.0 (2.0 ± 0.6)*	11.6 ± 2.4 (2.4 ± 0.7)*	31.7 ± 4.6	13.2 ± 6.5 (2.4 ± 1.2)*
Fgf4 element	42.5 ± 5.5	49.1 ± 5.0 (0.9 ± 0.1)*	25.2 ± 4.1 (1.7 ± 0.4)*	70.2 ± 19.1	23.3 ± 1.2 (3.0 ± 0.8)*
Utf1 element	32.0 ± 5.5	19.8 ± 7.3 (1.6 ± 0.7)*	15.5 ± 6.0 (2.1 ± 0.9)*	44.0 ± 9.8	43.4 ± 1.5 (1.0 ± 0.2)*

Table 1: aK_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.

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

Tapan Kumar Mistri et al. 2018

Table of contents:

FLUORESCENCE CORRELATION SPECTROSCOPY (THEORY)

FCS is a single molecule sensitive technique that provides quantitative information on molecular diffusion and concentration of fluorescent molecules within the sample. Fluorescence tagged molecules which diffuse through the confocal volume by Brownian motion generate fluorescence fluctuations. This fluctuation is converted into an autocorrelation curve by a normalized autocorrelation function (ACF):

$$G(\tau) = \frac{\langle F(t)F(t+\tau)\rangle}{\langle F(t)\rangle^2}$$

Equation (1)

In a binary binding experiment, where a fraction of the labelled molecules is unbound, (free) thereby exhibiting faster diffusion characteristics than a bound molecule in a binding complex, the experimental ACF is then fitted by a theoretical ACF describing two species with different diffusion times τ_D undergoing 3D free diffusion [1].

$$G(\tau) = \frac{1}{N}\left((1-y)\left(1+\frac{\tau}{\tau_{free}}\right)^{-1}\left[1+\left(\frac{\omega_0}{z_0}\right)^2\frac{\tau}{\tau_{free}}\right]^{-1/2} + y\left(1+\frac{\tau}{\tau_{bound}}\right)^{-1}\left[1+\left(\frac{\omega_0}{z_0}\right)^2\frac{\tau}{\tau_{bound}}\right]^{-1/2}\right) + G(\infty)$$

Equation (2)

Where N is the total number of molecules undergoing the different diffusion in the observation volume; ω_0 and z_0 are the radial and axial distances where the excitation intensity reaches $1/e^2$ of its value from the centre of the observation volume; τ_D is the diffusion time of the fluorophore through the observation volume; $G(\infty)$ is the convergence value of the autocorrelation curves for long times with the expected value of 1; *y* is the fraction of the bound species undergoing diffusion time of τ_{Di}.

The diffusion time τ_D reflects the size of the fluorescent tagged molecule, and in the case of binding, the size of the complex. It is related to the diffusion coefficient *D* by

$$\tau_D = \frac{{\omega_0}^2}{4D}$$

Equation (3)

The effective observation volume V_{eff} can be determined by using a fluorescent dye of known D, by first obtaining ω_0 followed by z_0 (from Equation (2) and (3)):

$$V_{eff} = \pi^{\frac{3}{2}} \omega_0^{\,2} \omega_z$$

Equation (4)

Lastly, the absolute concentration (in molar unit (M)) can be obtained by:

$$C = \frac{N}{N_A V_{eff}}$$

Equation (5)

N is the total number of fluorophores in the effective volume, V_{eff}, N_A is the Avogadro constant which is 6.022×10^{23}.

MATERIALS AND METHODS

Fusion Protein Design and Expression Vector Construction – Mouse Oct4 and Sox2 proteins were fused at their amino termini with the fluorescent proteins GFP and mCherry respectively to generate GFP-Oct4 and mCherry-Sox2 fusion proteins. Two rounds of PCR using the Expand High Fidelity PCR System (Roche) were performed to obtain the fusion construct**.** An initial round was performed for each protein separately using an outer primer to append the sites for NotI or SacI and an inner primer to append the linker sequence, encoding Gly-Gly-Ser-Gly, and a portion of nucleotide sequence encoding a part of the partner protein. A similar procedure was employed to make the Oct4-GFP and the Sox2-mCherry C-terminal fusions. The primers used for the initial PCR and Fusion PCR are shown below;

Plasmid (pCAG backbone)	**Primers**	**5'-3'**
pCAG GFP-Oct4 (N terminal fusion)	1F GFP	AATTGAGCTCTATGGCCACAACCATGGTGAGCAAGGGCGAGGAGC
	1R GFP	GAAGCCAGGTGTCCAGCCATGCCGCTGCCGCCCTTGTACAGCTCGTCCATGCC
	2F Oct4	GGCATGGACGAGCTGTACAAGGGCGGCAGCGGCATGGCTGGACACCTGGCTTC
	2R Oct4	AGTCGCGGCCGCTTCAGTTTGAATGCATGGGAG
pCAG Oct4-GFP (C terminal fusion)	1F Oct4	AATTGAGCTCCCGTCCCTAGGTGAGCCG
	1R Oct4	CCTCGCCCTTGCTCACCATGCCGCTGCCGCCGTTTGAATGCATGGGAGAGC

	2F GFP	GCTCTCCCATGCATTCAAACGGCGGCAGCGGCATGGTGAGCAAGGGCGAGG
	2R GFP	AGTCGCGGCCGCTTTACTTGTACAGCTCGTCC
pCAG mCherry-Sox2 (N terminal fusion)	1F mCherry	AATTGAGCTCTATGGCCACAACCATGGTGAGC
	1R mCherry	GCTTCAGCTCCGTCTCCATCATGCCGCTGCCGCCCTTGTACAGCTCGTCCATGC
	2F Sox2	GCATGGACGAGCTGTACAAGGGCGGCAGCGGCATGATGGAGACGGAGCTGAAGC
	2R Sox2	AGTCGCGGCCGCTTCACATGTGCGACAGGGGCAG
pCAG Sox2-mCherry (C terminal fusion)	1F Sox2	AATTGAGCTCCCAGCGCCCGCATGTATAACATG
	1R Sox2	CCTCCTCGCCCTTGCTCACCATGCCGCTGCCGCCCATGTGCGACAGGGGCAGTG
	2F mCherry	CACTGCCCCTGTCGCACATGGGCGGCAGCGGCATGGTGAGCAAGGGCGAGGAGG
	2R mCherry	AGTCGCGGCCGCTTTACTTGTACAGCTCGTCCATGC

Five fusion proteins were designed as follows:

1. Oct4 fused in its C-terminal with GFP (pCAG-Oct4-GFP-IN)
2. Oct4 fused in its N-terminal with GFP (pCAG-GFP-Oct4-IN)
3. Sox2 fused in its C-terminal with mCherry (pCAG-Sox2-mCherry-IN)
4. Sox2 fused in its N-terminal with mCherry (pCAG-mCherry-Sox2-IN)
5. mCherry fused in its C-terminal with GFP (pCAG-mCherry-GFP-IN), for FRET as Positive control

All fusion proteins are cloned into a pCAG vector to drive its expression in mouse embryonic stem cells.

These below three vectors were constructed without fusion PCR for FRET experiment.

1. GFP (pCAG-GFP-IN), for GFP
2. mCherry (pCAG-mCherry-IN), for mCherry
3. GFP-IRES-mCherry (pCAG-GFP-IRES-mCherry-IN), where IRES (Internal Ribosome Entry Site) allows GFP and mCherry to be expressed separately.

[NOTE: In all cases, IN stands for IRES-Neomycin cassette attached next to the protein of interest. In addition, GFP protein, used in this study, contains a mutation (alanine to lysine at amino acid 206) [2, 3] for reducing its tendency to form homodimers.]

Cell culture and transfection – ES cells were grown on a layer of mouse embryonic fibroblast in Dulbecco's modified Eagle's medium (DMEM, GIBCO-19600), 20% ES standard fetal bovine serum (FBS), 1x non-essential amino acids, 2 mM L-glutamine, 1X penicillin/streptomycin 0.1 mM 2-mercaptoethanol, and an aliquot of recombinant LIF conditioned medium. CHO-K1 cells (ATCC # CRL-61) were grown in Dulbecco's modified Eagle's medium (DMEM, GIBCO-19600), 10 % fetal bovine serum (GIBCO), 1x non-essential amino acids, 2 mM L-glutamine, 1X penicillin/streptomycin. The incubator conditions were set to 37 ºC, 5 % CO_2 and 95 % humidity. Transfections of required plasmids were performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. 24 µg of plasmid was transfected into CHO cells on 59 cm^2 plates. The cells were collected after 24 hours to prepare nuclear cell lysates.

Nuclear lysate preparation Nuclear extracts were prepared as described [4, 5].

Western blotting-Western blotting was performed using nuclear cell lysate (E14 mES and CHO cells line). DNA plasmids encoding the required TFs were transfected using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Cells were harvested after 24 hrs and nuclear lysates prepared. The blot was run following a standard protocol. Sox2 and Oct4 fusion proteins were detected using the following antibodies: for Oct4 Oct4-N19 primary (obtained from goat, Santa Cruz Biotechnology) (0.5 µl in 10 ml 5% blocker solution), for Sox2 Sox2-Y17 primary (obtained from goat, Santa Cruz Biotechnology) (2 µl in 10 ml %5 blocker solution). HPR conjugated rabbit anti-goat antibody (Santa Cruz Biotechnology) was used as secondary antibody in both cases. Signal development was done by incubating in super signal® west pico chemo luminescent substrate mixture [stable peroxide solution and luminal enhancer solution] (Pierce cat. # 34077) for 5 minutes. The final image was obtained using VersaDoc instrument (BioRad).

BINDING ASSAY FOR TITRATION

BINDING REACTIONS FOR TITRATION USED IN FCS AND EMSA									
In the absence of co-factor	Volume (μl)								
Protein (EO or MS)	0.0	0.5	1.0	1.5	2.0	3.0	5.0	7.0	9.0
BBC (make up to 9ul)	9.0	8.5	8.0	7.5	7.0	6.0	4.0	2.0	0.0
Cy5-NSO (150nM)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
BB (80% glycerol)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
dGdC (2ug/ul)	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Total volume	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0

In the presence of co-factor	Volume (μl)								
Protein (EO or MS)	0.0	0.5	1.0	1.5	2.0	3.0	5.0	7.0	8.0
BBC (make up to 9ul)	7.0	6.5	6.0	5.5	5.0	4.0	2.0	1.0	0.0
Cofactor (MS or EO)	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
Cy5-NSO (150nM)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
BB (80% glycerol)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
dGdC (2ug/ul)	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Total volume	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0

EO: GFP-Oct4, MS: mCherry-Sox2, BBC: Binding Buffer C, BB: Binding Buffer.

OLIGO SEQUENCES FOR QUANTITATIVE ASSAYS

Name (motif)	Sequence	Figure
Nanog -SO	TCCACCATGGA**CATTGTAATGCAAA A**GAAGCTGTAAG	Fig. 3, Fig. S3, Table-1
Utf1-SO	GGATGAGCCGT**CATTGTTATGCTAG T**GAAGTGCCGGC	Fig. S3, Table-1
Fgf4-SO	TAGAAAACT**CTTTGTTTGGATGCTA AT**GGGATACTTA	Fig. 4, Fig. 5, Fig. S3, Table-1
Oct4-SO	CTATCATGCAC**CTTTGTTATGCATC T**GCCGTCTGCCC	Fig. S3
Sox2-SO	CTCGGGCAGC**CATTGTGATGCATAT** AGGATTATTCA	Fig. S3
Fgfr2-S	GCCACCCCACTTCATTGTAGCATTTC TTTCATTTGTC	Fig. 5

Supplementary Table 1: Oligonucleotides used for titration assays. DNAs used are shown 5'-3'. In red is the Sox2 binding site and in green is that of. All the motifs are labeled with Cy5.

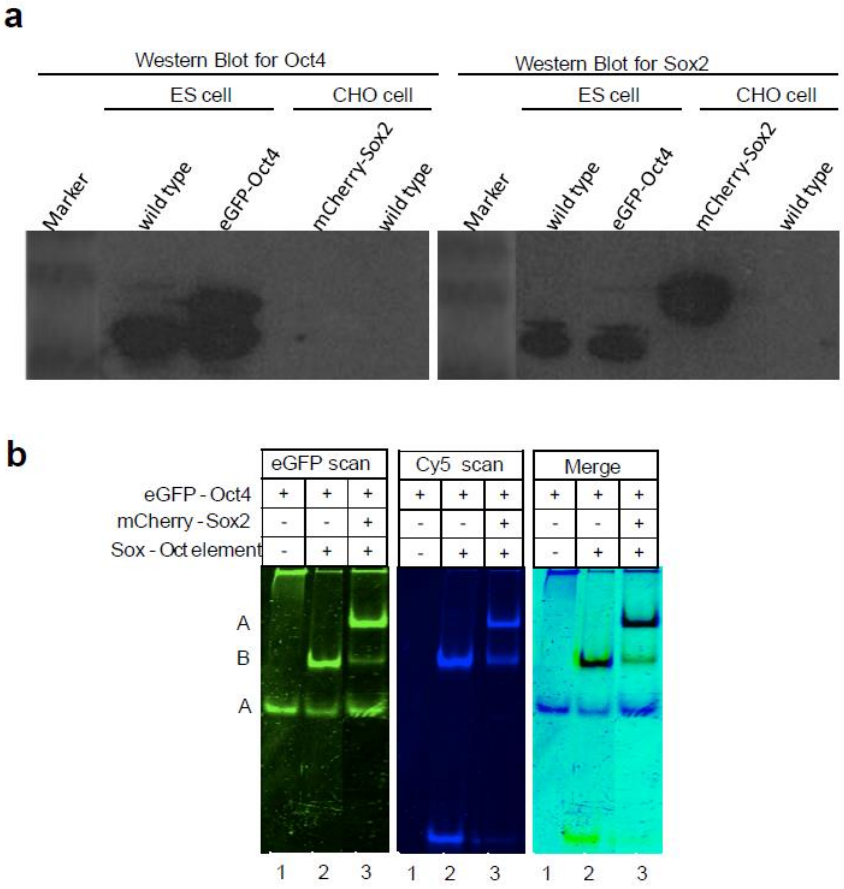
OLIGO SEQUENCES FOR FP-EMSA

Name (motif)	Sequence	Figure
NSO	GGACATTGTAATGCAAAAGAA	Fig. S3a-b
NSO_mut	GGACAaataAgcctAAAAGAA	
NSO _mO5-8C	GGACATTGTAATGCccccGAA	Fig. S3a
NSO _mO5-8T	GGACATTGTAATGCttttGAA	
NSO _mO5-8G	GGACATTGTAATGCggggGAA	
NSO _mO1-4A	GGACATTGTAAaaaAAAAGAA	
NSO _mO1-4G	GGACATTGTAggGgAAAAGAA	
NSO _mO1-4C	GGACATTGTAcccCAAAAGAA	
NSO _mO1-4T	GGACATTGTAtTttAAAAGAA	
NSO_O1	CATGGACATTGTAcTGCAAAAGAAGCT	Fig. S3b
NSO_O2	CATGGACATTGTAAaGCAAAAGAAGCT	
NSO_O3	CATGGACATTGTAATaCAAAAGAAGCT	
NSO_O4	CATGGACATTGTAATGgAAAAGAAGCT	
NSO_O1-2	CATGGACATTGTAcaGCAAAAGAAGCT	
NSO_O3-4	CATGGACATTGTAATagAAAAGAAGCT	
NSO_O1-4	CATGGACATTGTAcaagAAAAGAAGCT	

Supplementary Table 2: Oligonucleotides for testing sequence specificity of Oct4 binding to *Nanog* Sox/Oct (NSO) motif. DNAs used are shown 5'-3'. In red is the Sox2 binding site and in green is that of Oct4. Mutations are marked by lower case (substitutions). All the motifs here are unlabeled (unless otherwise mentioned in the Figure legend).

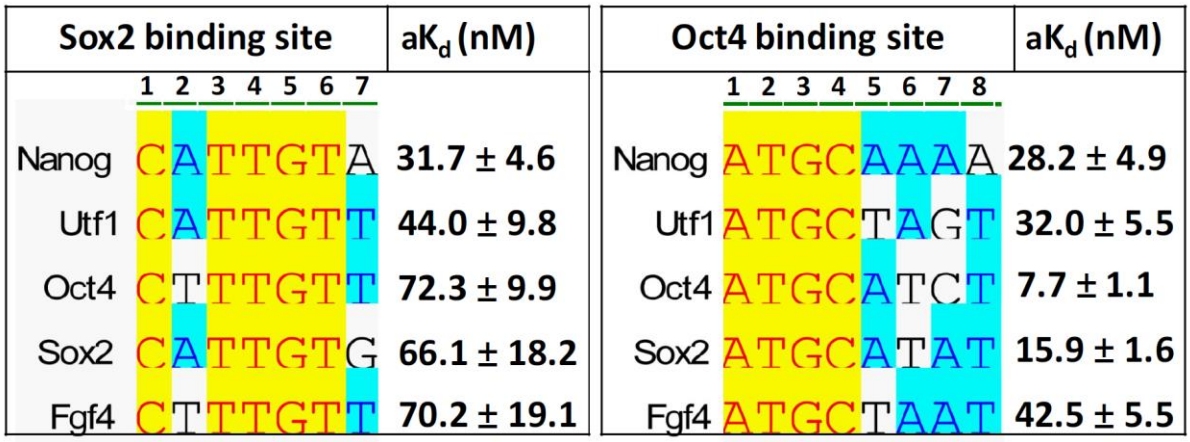
SUPPLEMENTARY FIGURES

SUPPLEMENTARY FIGURE 1.



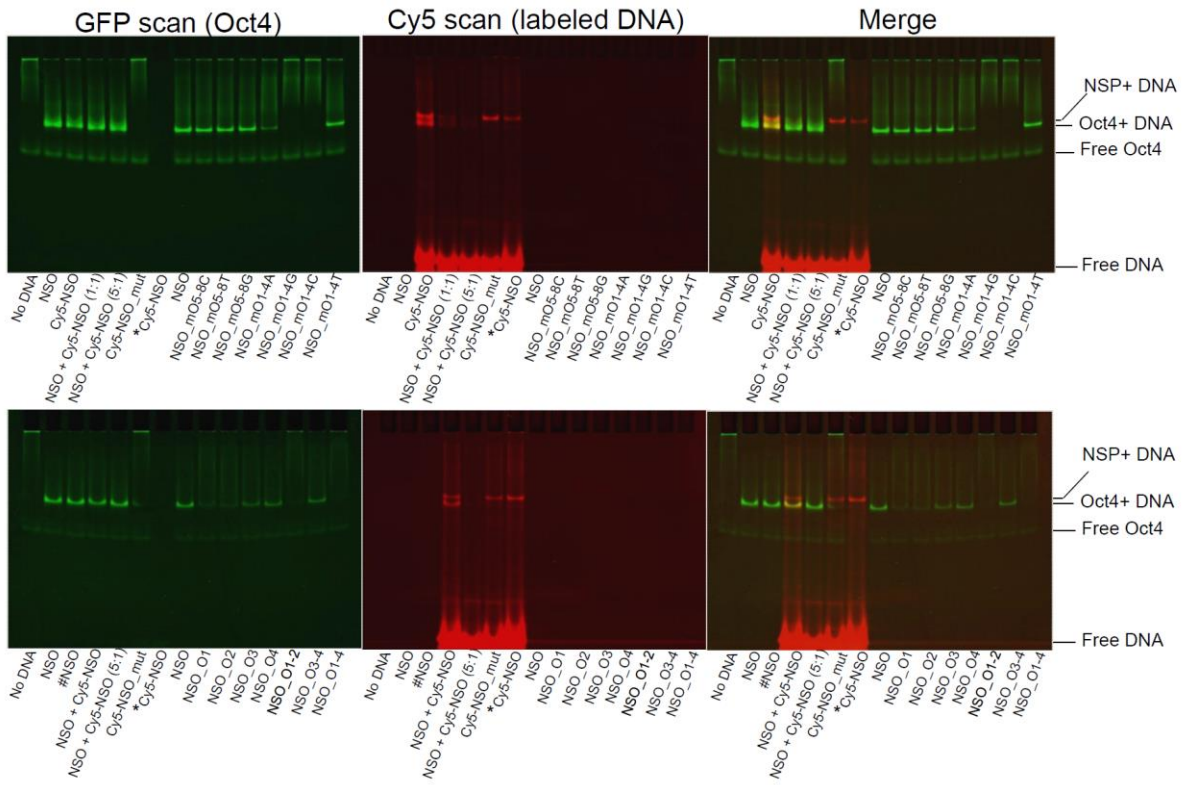
SUPPLEMENTARY FIGURE 1. **CHO cell line is ideal for fusion protein generation.** (**a**) Western blot results for Oct4 (left) and Sox2 (right) in ES or CHO nuclear cell lysates. R = 3. (**b**) Complex formation of GFP-Oct4 with the *Nanog* Sox/Oct motif in the presence or absence of mCherry-Sox2 by FP-EMSA. R = 3.

SUPPLEMENTARY FIGURE 2.



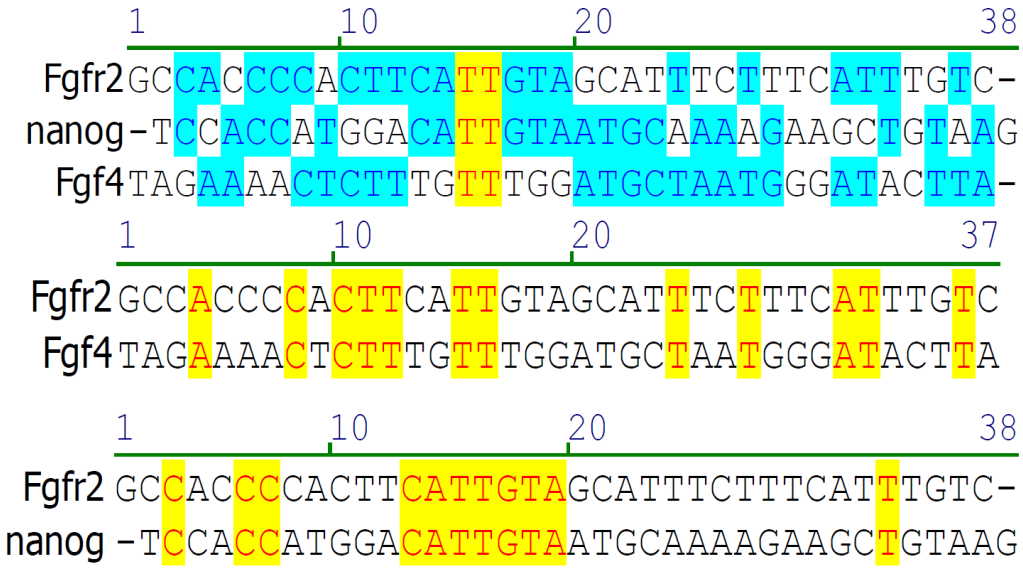
SUPPLEMENTARY FIGURE 2. **aK_d of Sox/Oct motifs measured for either Oct4 or Sox2**. The indicated Sox/Oct motifs were compared to show the Sox2 binding site sequence (left) and the Oct4 binding site sequence (right) as well as aK_d values measured for either Sox2 or Oct4 alone by FCS. R = 3, mean ± SEM.

SUPPLEMENTARY FIGURE 3.



SUPPLEMENTARY FIGURE 3. **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**. R = 2

SUPPLEMENTARY FIGURE 4.



SUPPLEMENTARY FIGURE 4. **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.

REFERENCE

1 Kinjo, M. and Rigler, R. (1995) Ultrasensitive hybridization analysis using fluorescence correlation spectroscopy. Nucleic acids research. **23**, 1795

2 Shaner, N. C., Steinbach, P. A. and Tsien, R. Y. (2005) A guide to choosing fluorescent proteins. Nat Methods. **2**, 905-909

3 Zacharias, D. A., Violin, J. D., Newton, A. C. and Tsien, R. Y. (2002) Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. Science. **296**, 913

4 Andrews, N. C. and Faller, D. V. (1991) A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. Nucleic acids research. **19**, 2499

5 Mistri, T. K., Devasia, A. G., Chu, L. T., Ng, W. P., Halbritter, F., Colby, D., Martynoga, B., Tomlinson, S. R., Chambers, I. and Robson, P. (2015) Selective influence of Sox2 on POU transcription factor binding in embryonic and neural stem cells. EMBO reports. **16**, 1177-1191