Emerging cooperativity between Oct4 and Sox2 governs the pluripotency network in mouse embryos

Yanlin Hou[1,2], Zhengwen Nie[2], Sergiy Velychko[1,3], Ivan Bedzhov[1], Sandra Heising[1], Qi Jiang[2], Hui Zhang[2], Guangming Wu[2 *], Kenjiro Adachi[1,4 *], Hans R. Schöler[1,5 *]

[1] Max Planck Institute for Molecular Biomedicine, Röntgenstrasse 20, 48149 Münster, Germany

[2] Guangzhou National Laboratory, 9 XingDaoHuanBei Road, Guangzhou International Bio Island, Guangzhou 510005, Guangdong, China

[3] Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

[4] Present address: Astellas Institute for Regenerative Medicine, Westborough, MA 01581, USA

[5] Lead Contact

[*] Correspondence: office@mpi-muenster.mpg.de, kenjiro.adachi@astellas.com, wu_guangming@gzlab.ac.cn

Abstract

During the first lineage segregation, mammalian embryos generate the inner cell mass (ICM) and trophectoderm (TE). ICM gives rise to the epiblast (EPI) that forms all cell types of the body, an ability referred to as pluripotency. The molecular mechanisms that induce pluripotency in embryos remain incompletely elucidated. Using knockout (KO) mouse models in conjunction with low-input ATAC-seq and RNA-seq, we found that Oct4 and Sox2 are globally dispensable in morulae. Oct4 and Sox2 come into play in the early ICM, coinciding with the initiation of Sox2 expression, and they directly activate the pluripotency-related genes and the corresponding OCT-SOX enhancers. *Sox2* KO failed to activate the ICM-specific pluripotency-related genes while overexpression of Sox2 prematurely upregulated these genes in morulae, suggesting that Sox2 acts as a limiting factor for inducing the pluripotent network. Our study provides new insights into this critical process of mouse preimplantation development at molecular levels.

Keywords chromatin accessibility, gene transcription, inner cell mass, Oct4, Sox2

Introduction

Following fertilization, mouse embryos undergo a series of cell divisions and differentiation in the first few days. Zygotes and 2-cell blastomeres are totipotent and can form both embryonic and extra-embryonic supporting tissues. After several rounds of divisions, embryos undergo compaction at the morula stage and specifies the trophectoderm (TE) and inner cell mass (ICM). The ICM further segregates into the primitive endoderm (PE) and the epiblast (EPI) in the subsequent blastocyst stage. All cell types of the fetus can develop from the EPI, a feature known as pluripotency. During the establishment of pluripotency, dramatic molecular changes occur such as shifts in epigenetic modifications, chromatin accessibility and transcriptome (Deng *et al*, 2014; Guo *et al*, 2010; Wang *et al*, 2018; Wu *et al*, 2016; Zhang *et al*, 2016; Zhang *et al*, 2018). The precise mechanisms involved in establishing the transient pluripotent state during development remain elusive.

Oct4 (also known as *Pou5f1*) and Sox2 are prominent transcription factors (TFs) that regulate early embryonic development. In mice, Oct4 is expressed in all the cells of the compacted morulae, while Sox2 is first expressed in the inside cell and identified as one of the earliest markers to distinguish the inner from the outer cells (Guo *et al.*, 2010; Palmieri *et al*, 1994; White *et al*, 2016). Both factors are confined to the pluripotent EPI at the late blastocyst stage (Avilion *et al*, 2003; Palmieri *et al.*, 1994). *Oct4* and *Sox2* knockout (KO) leads to embryonic lethality around embryonic day (E) 4.5 and E6.0, respectively (Avilion *et al.*, 2003; Nichols *et al*, 1998). Although *Oct4*- or *Sox2*-KO embryos develop apparently normal EPI, they fail to give rise to embryonic stem cells (ESCs) (Avilion *et al.*, 2003; Nichols *et al.*, 1998) and *Oct4*-KO embryos cannot contribute to embryonic tissues in chimera assays (Wu *et al*, 2013), suggesting that Oct4 and Sox2 are required for the EPI to acquire or maintain pluripotency.

Controversially, *Oct4*- and *Sox2-KO* mouse embryos still express many pluripotency markers, such as Nanog, *Klf4* and *Zfp42* (*Rex1*) (Avilion *et al.*, 2003; Frum *et al*, 2013; Le Bin *et al*, 2014; Stirparo *et al*, 2021; Wicklow *et al*, 2014; Wu *et al.*, 2013), suggesting that the pluripotency network is maintained. An important question to be addressed is whether Oct4 and Sox2 are required for the establishment of the pluripotency gene network.

Recent studies suggested that Oct4 and Sox2 may function before the blastocyst stage. The long-lived presence of exogenous Oct4 and Sox2 proteins on DNA was observed in some of the 4-cell blastomeres, that tended toward the ICM rather than the TE (Plachta *et al*, 2011; White *et al.*, 2016), suggesting that they may play a role in the initial lineage segregation between ICM and TE. Oct4 is also thought to contribute to the dramatic gain of chromatin accessibility at the 8-cell stage (Lu *et al*, 2016). Unlike in mouse embryos, the loss of OCT4 expression in human embryos compromises blastocyst development, leading to the downregulation of *NANOG* in the ICM and *CDX2* in the TE (Fogarty *et al*, 2017). Oct4 controls zygotic genome activation (ZGA) in zebrafish and human, but not in mice (Gao *et al*, 2018; Leichsenring *et al*, 2013). Thus, when and how Oct4 and Sox2 function in the early embryogenesis varies from species to species.

Oct4 and Sox2 play preeminent roles in maintaining the pluripotency network in ESCs (Adachi *et al*, 2013; Masui *et al*, 2007; Niwa *et al*, 2000). However, it remains unclear how these two TFs regulate pluripotency genes in embryos. The obvious obstacles are the heterogeneity and the scarcity of the preimplantation embryos. Unlike ESCs, which comprise a homogenous cell population, individual cells within embryos gradually differentiate into

three cell types, EPI, PE and TE. This inherent cellular heterogeneity within embryos, coupled with the temporal variability in development across different embryos, poses a significant challenge when investigating the molecular processes using conventional molecular techniques. For instance, using the entire blastocyst or ICM for qPCR and bulk RNA-seq analysis would inevitably lead to a dilution and blur of the molecular processes within EPI (Frum *et al.*, 2013; Wu *et al.*, 2013). Although single-cell genomics is powerful to dissect cellular heterogeneity, it suffers from high dropouts, which may lead to inaccuracies in genomic profiles and cell identities (Minow *et al*, 2023). Due to the scarcity of embryos, it is unfeasible to compensate for the inaccuracy by increasing the number of cells.

MEK inhibitor (MEKi) promotes the ground state of ESCs by suppressing the PE-inducing ERK signaling pathway (Ying *et al*, 2008). In embryos, MEKi has been shown to activate Nanog in ICMs, effectively suppressing the development of PE and directing the entire ICM towards a pluripotent epiblast fate (Nichols *et al*, 2009). Naïve pluripotency in MEKi-treated EPI has been confirmed by its contribution to chimaeras with germline transmission. MEKi may thus facilitate the study of molecular events within the EPI cells by reducing the heterogeneity.

In this study, we analyzed the transcriptome and global chromatin accessibility in the *Oct4*- or *Sox2*-KO mouse embryos using low-input RNA-seq and ATAC-seq. The effects of *Oct4* and *Sox2* KO are small in morulae, but become evident in ICMs at the blastocyst stage. Oct4 and Sox2 mainly activate pluripotency-related genes cooperatively through the enhancers containing the composite OCT-SOX motifs. At the same time, Oct4 can weakly maintain the accessibility of OCT-SOX enhancers in the absence of Sox2. We observed that *Sox2* KO caused downregulation of the ICM-specific pluripotency-related genes, while forced expression of Sox2 could prematurely induce these genes in morulae, suggesting that the onset of Sox2 expression induces the activation of the pluripotency network in mouse preimplantation embryos.

Results

Generation of maternal-zygotic KO embryos

To investigate the role of Oct4 and Sox2 in the development of preimplantation embryos, we produced four transgenic mouse lines: *Oct4* KO labeled with mKO2 (monomeric kusabira-orange 2, $Oct4^{mKO2}$), *Sox2* KO labeled with EGFP ($Sox2^{EGFP}$), floxed *Oct4* ($Oct4^{flox}$) and floxed *Sox2* ($Sox2^{flox}$) (Fig 1A, 1B; methods). Although the expression of mKO2 and EGFP is driven by the constitutive mouse PGK promoter, we observed that the expression of these reporters is weak until the 8-cell and blastocyst stages, respectively, when the strong zygotic expression of Oct4 and Sox2 is observed.

$Oct4^{+/mKO2}$ and $Sox2^{+/EGFP}$ heterozygous male mice were mated with $Oct4^{flox/flox}$; *ZP3-Cre* and $Sox2^{flox/flox}$; *ZP3-Cre* maternal KO female mice, respectively, to generate fluorescently labeled maternal-zygotic KO and unlabeled maternal KO (herein called control, Ctrl) embryos (Fig 1C, 1D). These reporter systems allow prospective identification of KO embryos using fluorescent microscopy (Appendix Fig S1), so we were able to pool embryos with the same genotype. Immunostaining results confirmed that the mKO2+ and EGFP+ embryos lost Oct4 and Sox2 proteins, respectively (Fig 1E, 1F).

Oct4 and Sox2 regulate the chromatin landscape and transcriptome in the ICM, but not in the morula

To assess the genome-wide molecular impact of the loss of Oct4 and Sox2, we performed low-input ATAC-seq and RNA-seq using early (E2.75) and late (E3.25) compacted morulae, as well as early (E3.75) and late (E4.5) ICMs (Methods). For ATAC-seq, we pooled embryos based on our reporter systems because single-embryo samples yielded sparse signals and too few peaks (data not shown). We opted to collect single morulae or ICMs for RNA-seq, as this approach enabled us to account for embryo-to-embryo variability and detect transcripts with greater sensitivity compared to scRNA-seq (Boroviak *et al*, 2015). Given that the ICM consists of the progenitors of EPI and PE cells, we treated the embryos with the MEKi (PD0325901) from E2.5 to suppress PE development and specify the entire ICM to the EPI (Nichols *et al.*, 2009) (Appendix Fig S2).

Principal component analysis (PCA) of the ATAC-seq data shows the samples clustered according to the genotypes and stages (Fig 2A). The *Oct4*-KO early and late morulae clustered close to their Ctrl counterparts, while the *Oct4*- and *Sox2*-KO early and late ICM samples clustered separately from their Ctrl. k-means clustering shows that the dynamics of chromatin accessibility between different stages and genotypes were captured (Fig 2B). Many of the peaks that exhibited reduced accessibility in the KO ICMs gained accessibility during morula to ICM transition (clusters 3, 5, 6, 10, 17, 19 and 27). Notably, the peaks activated by Oct4 but not by Sox2 tended to be open in morulae (clusters 13, 17, and 27), whereas those activated by both Oct4 and Sox2 became open in ICMs (clusters 3, 10, and 19). To quantify the impacts of *Oct4* or *Sox2* deletion on the chromatin accessibility, we used the significantly increased and decreased peaks by DESeq2 (Fig 2C, Supplementary data 1, methods). Out of the 152,393 detected peaks, 14,016 (9.2%) and 6,637 (4.4%) showed significant decreases, while 11,884 (7.8%) and 3,660 (2.4%) exhibited significant increases in *Oct4* and *Sox2* KO late ICM, respectively. The number of decreased peaks exceeded that of increased peaks across all the KO samples. Over 96.8% of the significantly changed peaks were identified as putative enhancers located distally to the transcription start sites (TSSs) (Appendix Fig S3A). Analysis of published ChIP-seq data in ESCs (Marson *et al*, 2008; Whyte *et al*, 2012) shows that enhancers decreased in the *Oct4*- or *Sox2*-KO early ICMs were frequently bound by Oct4, Sox2 and Nanog, and enriched with the active enhancer mark H3K27ac (Appendix Fig S3B), suggesting that they could be the direct targets of Oct4 and Sox2 in the ICM as well.

To assess the impact of *Oct4* and *Sox2* deletion on the transcriptome, we performed RNA-seq on individual morulae or ICMs. Initially, we examined the consequences of maternal KO and zygotic heterozygous KO. PCA did not reveal significant distinctions between maternal flox and maternal KO, nor between WT/flox and flox/KO samples (Appendix Fig S3C). Furthermore, mice with heterozygous KO of *Oct4* or *Sox2* develop normally and exhibit fertility (Nichols *et al.*, 1998). Therefore, we employed maternal KO zygotic WT/KO as our control (Ctrl) group for the analyses (Fig 1C, 1D, S3C). *Oct4* or *Sox2* KO altered the expression of 2,485 (15.8%) and 967 (6.1%) genes in the late ICM, respectively (Fig 2D, Supplementary data 2). Interestingly, *Oct4* KO had only a minor effect on the transcriptome (132, 0.8%) in morulae, indicating that Oct4 plays a limited role despite its high expression at this stage. As Sox2 is only expressed at very low levels until the blastocyst stage anyway, the transcriptome in *Sox2* KO morulae was not significantly affected. *Oct4* KO had a greater effect than *Sox2* KO on both the chromatin accessibility and the transcriptome (Fig 2A-2D),

suggesting Oct4 may play a more important role than Sox2. This is consistent with the earlier developmental arrest in *Oct4*-KO (E4.5) than in *Sox2*-KO (E6.0) embryos.

Next, we investigated if the change in chromatin accessibility affected the transcription of the surrounding genes. Gene set enrichment analysis (GSEA) shows that the genes proximal to the decreased peaks were frequently downregulated, whereas those near the increased peaks were upregulated in *Oct4*- and *Sox2*-KO embryos (Fig 3A-3D). For example, the expression of *Utf1*, a known target of Oct4 and Sox2, and its enhancer accessibility decreased in the *Oct4*- and *Sox2*-KO ICMs (Fig 3E, 3F). On the other hand, the expression of *Fgfr2* and its enhancer accessibility increased in the *Oct4*- and *Sox2*-KO ICM. In addition, integration of the ATAC-seq and RNA-seq data allowed us to identify previously unknown targets of Oct4 and Sox2, such as *Sap30* and *Uhrf1*. Sap30 is a key component of the mSin3A corepressor complex, which facilitates silencing of somatic genes during reprogramming (Li *et al*, 2017).

Taken together, the data presented so far indicate that Oct4 and Sox2 play a crucial role in shaping the chromatin landscape and transcriptome in the ICM, but not in the morula. Consequently, we observed apparently normal development of the *Oct4*- and *Sox2*-KO embryos up to the blastocyst stage.

Oct4 and Sox2 activate the pluripotency network in the ICM

To further explore the role of Oct4 and Sox2, we wanted to find out which genes and resulting aspects within the ICM they might influence. Oct4 and Sox2 are among the master TFs known to cooperatively maintain the pluripotency network in ESCs (Chen *et al*, 2008). As expected, many of the differential ATAC-seq peaks were consistently changed both in the *Oct4*- and *Sox2*-KO late ICM (Fig 4A, Supplementary data 1). However, there are a number of peaks exclusively changed in either *Oct4*- or *Sox2*-KO ICMs. The genes around the consistently decreased peaks were enriched for the gene ontology (GO) terms related to pluripotency and preimplantation embryonic development, such as cellular response to LIF signaling and stem cell population maintenance (Fig 4B, upper panel). The consistently elevated peaks preferentially located near the genes related to extraembryonic lineages and organogenesis, such as embryonic placenta, extraembryonic trophoblast, muscle and neural tube (Fig 4B, lower panel). Interestingly, the genes near the peaks which decreased only in the *Oct4*- but not *Sox2*-KO ICM were enriched with terms of LIF signaling and blastocyst formation (Appendix Fig S4).

At the transcriptional level, many known pluripotent genes, such as *Klf2, Etv5, Prdm14* and *Pecam1*, were downregulated in the *Oct4*- and *Sox2*-KO ICM (Fig 4C). In contrast, *Gata3, Cdx2* and *Eomes*, which are important for TE development and differentiation (Ralston *et al*, 2010; Russ *et al*, 2000; Strumpf *et al*, 2005), were upregulated in the *Oct4*- and Sox2-KO ICM. Interestingly, we observed that *Nanog*, *Esrrb* and *Klf4* were significantly downregulated in the *Oct4*-KO ICM, whereas in the *Sox2*-KO ICM they were not or only slightly downregulated (Fig 4C, 4D). Downregulation of *Pecam1* was confirmed at the protein level (Fig 4E). Chromatin accessibility at its putative enhancers also decreased accordingly (Fig 4F). Oct4 and Sox2 activated the components of several epigenetic modifiers, including *Ezh2* (PRC2 H3K27me2/3 methyltransferase) and *Sap30* (mSin3A histone deacetylase complex) (Fig 3E, 3F, Supplementary data 2). Therefore, Oct4 and Sox2 may contribute to establishing the ICM-specific epigenetic status through direct regulation of the epigenetic modifiers.

Taken together, the above data show that Oct4 and Sox2 regulate a large number of TFs, epigenetic factors and signaling pathways in the ICM. Furthermore, Oct4 can maintain a subset of TFs in the naïve pluripotency regulatory circuit independently of Sox2.

OCT-SOX enhancers are co-activated by Oct4 and Sox2, with Oct4 playing a dominant role

Although there is a general consensus on the cooperative binding of Oct4 and Sox2 to the OCT-SOX composite motif, the principle of the cooperation still remains controversial in different scenarios (Biddle *et al*, 2019; Chen *et al*, 2014; Friman *et al*, 2019; Li *et al*, 2019; Michael *et al*, 2020). We investigated how Oct4 and Sox2 regulate their target enhancers in the ICM. As expected, the OCT-SOX, OCT and SOX motifs were also among the most enriched motifs in the group of decreased enhancers in *Oct4*- or *Sox2*-KO ICMs (Fig 5A), suggesting that MEKi did not alter the direct target enhancers of Oct4 and Sox2. Motifs of other known pluripotency-related TFs, such as Klf2/4 and Esrrb, were enriched at the decreased enhancers. The elevated peaks were enriched for the GATA, TEAD, EOMES and KLF motifs, but not for the OCT-SOX, OCT or SOX motifs. Thus, the increased peaks may not be the direct targets of Oct4 or Sox2 but likely the secondary effect of the upregulated TE genes, such as *Gata3*, *Tead4*, *Eomes* and *Klf5/6* (Fig 4C, Supplementary data 2). As a control, the CTCF motif was not enriched in either the decreased or increased peaks. We confirmed the enrichment of those motifs at distal peaks and concluded that the motif analysis was not biased by the small number of the TSS peaks in the dataset (Appendix Fig S3A and S5).

To investigate the effect of Oct4 and Sox2 on the chromatin accessibility, we focused on the peaks in the ICM. It is worth noting that the top decreased peaks were most enriched with the OCT-SOX, but not the OCT or SOX motifs, in the *Oct4*- and *Sox2*-KO early ICM (Fig 5B), which suggests that peaks containing the OCT-SOX motif (hereafter referred to as OCT-SOX peaks) might be the primary targets of Oct4 and Sox2. The KLF motif was enriched at the moderately decreased peaks, indicating a possible secondary effect of down-regulated *Klf2/4* (Fig 4C, 4D); alternatively, the depletion of Oct4 and Sox2 might lead to a decrease in DNA binding or the activity of Klf2/4 and other KLF TFs.

In general, the accessibility of 8,993 OCT-SOX peaks decreased in both the *Oct4*- and *Sox2*-KO early and late ICM (Fig 5C, 5D, S6), suggesting that both TFs participate in keeping these peaks open. Notably, the accessibility of the OCT-SOX peaks decreased more for *Oct4* KO than for *Sox2* KO. To explore the reasons behind the differential control of OCT-SOX enhancers by Oct4 and Sox2, we further categorized the OCT-SOX peaks based on whether they have mismatches (mm) on the highly conserved bases in the Oct4 POU-specific domain-binding site (GCAT) and Sox2-binding site (ACAA) (Fig 5E). The OCT-SOX peaks harboring conserved Sox2-binding sites (676 canonical OCT-SOX and 1,782 OCTmm-SOX) exhibited similar reductions in the *Oct4* KO and *Sox2* KO ICM (Fig 5F). On the contrary, the OCT-SOX peaks with mismatches in the Sox2-binding sites (3,297 OCT-SOXmm and 3,710 OCTmm-SOXmm) showed a more pronounced decrease in *Oct4*-KO than in *Sox2*-KO early ICMs. This suggests that the higher dependency of the OCT-SOX peaks on Oct4 can partly be attributed to the limited influence of Sox2 on those OCT-SOX peaks containing mismatches at the Sox2-binding site. The known OCT-SOXmm enhancers of *Esrrb* and *Klf4* are shown as examples (Fig 5G).

The above results suggest that the enhancers containing the OCT-SOX motif are primary targets of Oct4 and Sox2 in the ICM, and a subset of these enhancers is predominantly regulated by Oct4.

Sox2 acts as a limiting factor for the activation of EPI-lineage genes

In reprogramming of fibroblasts, activation of endogenous Sox2 is a late-stage and deterministic event that triggers the fate towards pluripotency (Buganim *et al*, 2012). While in embryos, the identity of blastomeres is not determined until the early blastocyst stage (Tarkowski *et al*, 2010). We then aimed to explore factors that determine the fate of pluripotent cells in the early ICM. In accordance with previous studies (Guo *et al.*, 2010; Wu *et al.*, 2016), we observed a substantial reorganization of open chromatin regions upon the formation of the early ICM (Fig 2B, 6A). Genes around the elevated peaks in ICM were involved in cellular response to LIF, maintenance of stem cell population and blastocyst formation (Appendix Fig S7A), consistent with the developmental stage. The top elevated peaks in ICM were enriched for the OCT, SOX and OCT-SOX motifs (Appendix Fig S7B). Furthermore, around 50% of the peaks and genes regulated by Oct4 and Sox2 were activated in the early ICM *vs* morula (Appendix Fig S7C). Above results indicate the involvement of Oct4 and Sox2 in the initiation of the ICM-specific chromatin and transcriptional programs.

In alignment with these chromatin dynamics, the transcriptome underwent significant rearrangements in early ICMs (Fig 6B). The network of naïve pluripotency is maintained by a core regulatory circuit of TFs, including Sox2, Oct4, Nanog, Esrrb and Klf4 (Adachi *et al*, 2018; Ng & Surani, 2011; Young, 2011). In the morula, all the aforementioned TFs except Sox2 are highly expressed (Fig 6C), their expression does not appear to be sufficient to induce the pluripotent state. Hence, we hypothesized that Sox2 may act as a limiting factor in the formation of the pluripotent EPI. To assess if this hypothesis is valid, we examined the impact of *Sox2* KO on the 2,462 upregulated genes in the early ICM *vs* morula (Fig 6B). Combined GSEA and Wikipathway analysis revealed a significant downregulation of the genes associated with the PluriNetWork and ESC pluripotency pathways upon *Sox2* KO in the early ICM (Fig 6D, 6E).

We then injected *Sox2* mRNA into the zygote cytoplasm to test if exogenous Sox2 could induce the premature onset of the pluripotency-related genes (Fig 6F). The Sox2 protein expression was validated by immunostaining of the morula 2.5 days later (Fig 6G). Of the 2,462 elevated genes in the early ICM, genes like *Fgf4*, *Etv5* and *Nanog* were upregulated in the Sox2-overexpressed morula (Fig 6B, 6H). In contrast, *Pou5f1*, *Esrrb* and *Klf4* were not significantly induced upon *Sox2* overexpression, consistent with their sustained expression between morulae and ICMs (Fig 6C). Furthermore, *Sox2* overexpression suppressed the TE marker *Cdx2*, *Gata3* and *Dab2*. These findings show that exogenous Sox2 can trigger premature activation of ICM-specific genes, highlighting the role of Sox2 as a limiting factor for the establishment of the pluripotency network in mouse embryos.

Discussion

Oct4 and Sox2 are key TFs for preimplantation embryonic development in several species (Daigneault *et al*, 2018; Fogarty *et al.*, 2017; Gao *et al*, 2022; Lee *et al*, 2013; Leichsenring *et al.*, 2013). However, because of the paucity of embryonic cells, it remains a matter of debate as when they start to function and how they regulate the development of the EPI

lineage. In this study, we explored the roles of Oct4 and Sox2 in mouse preimplantation embryos using transgenic mouse models.

Previous reports suggested that Oct4 and Sox2 could direct 4-cell blastomeres towards ICM fate at the first lineage segregation (Plachta *et al.*, 2011; White *et al.*, 2016). siRNA-mediated knockdown experiments showed that Oct4 facilitates chromatin opening already at the 8-cell stage (Lu *et al.*, 2016). Unexpectedly, our data showed that Oct4 and Sox2 have minimal impact on both global chromatin accessibility and transcription until the early blastocyst stage (Fig 2A-2D). Oct4 is expressed at similar levels between morulae and ICMs; in contrast, Sox2 mRNA and protein levels are negligible in morulae, but are significantly upregulated in ICMs (Fig 6C) (Palmieri *et al.*, 1994; Wicklow *et al.*, 2014). Therefore, the function of Oct4 is likely limited by the insufficient levels of Sox2 and possibly other factors in morulae. Another plausible limiting aspect could be epigenetic constraints in morulae. The epigenetic state and 3-dimensional structure of the chromatin in the morula differ from those in the ICM, which may hinder Oct4 binding (Du *et al*, 2017; Wang *et al.*, 2018; Wang *et al*, 2014; Zhang *et al.*, 2016).

Oct4 is activated prior to Sox2 in both embryogenesis and reprogramming to iPSCs. In mouse embryos, Oct4 is initially highly expressed in all blastomeres from the 8-cell stage and later becomes restricted to the EPI, whereas Sox2 is repressed by the Hippo pathway in early stages and is specifically expressed in the EPI (Frum *et al*, 2019; Guo *et al.*, 2010; Wicklow *et al.*, 2014). In reprogramming, the activation of endogenous Sox2 occurs later than the endogenous Oct4 and signifies the final phase of iPSC generation (Buganim *et al.*, 2012). In both scenarios, the expression of Sox2 coincides with the setup of the pluripotent state. Moreover, our data show that *Sox2* KO ICMs fail to fully activate the pluripotent program, whereas premature expression of Sox2 alone upregulates the EPI-specific genes in morulae (Fig 4C, 6H). Recently, we reported that the loss of Sox2 might contribute to decreased developmental potential of pluripotent cells upon priming (MacCarthy *et al*, 2022). This suggests that Sox2 plays a critical role in the spatial-temporal onset of the pluripotency-related program.

Loss of Oct4 and Sox2 impairs embryonic pluripotency, as evidenced by the inability of KO embryos to give rise to ESCs or contribute to the embryonic part in chimera assays (Avilion *et al.*, 2003; Nichols *et al.*, 1998; Wu *et al.*, 2013). To date, only a few of pluripotency-related genes have been found downstream of Oct4 and Sox2 in mouse embryos, whereas the majority, including *Nanog*, *Klf4* and *Zfp42* (*Rex1*), are reportedly independent of Oct4 and Sox2 (Frum *et al.*, 2013; Stirparo *et al.*, 2021; Wu *et al.*, 2013). Here, we showed that both Oct4 and Sox2 are essential for establishing the EPI-specific chromatin landscape and transcriptome (Fig 4A, 4C, 6D, S7B). We found a large number of pluripotency-related genes affected in *Oct4*-KO ICM, contrasting with previous reports suggesting that most pluripotency-related genes were maintained (Frum *et al.*, 2013; Le Bin *et al.*, 2014; Stirparo *et al.*, 2021; Wu *et al.*, 2013). This discrepancy may be due to differences in the experimental design. First, unlike previous study that used the whole blastocysts (Frum *et al.*, 2013), we managed to isolate a purer EPI cell population performing immunosurgery and using short-term inhibition of MEK to reduce the heterogeneity of the ICMs (Nichols *et al.*, 2009) (Fig S2). Second, we removed *Oct4* exons 2 to 5 (Fig 1), encompassing the whole POU domain (bipartite DNA-binding domain of Oct4), resulting in a complete ablation of Oct4 function. In contrast, some of the previous studies removed only exon 1 (Frum *et al.*, 2013; Wu *et al.*, 2013), which encodes for N-terminal transactivation domain. The remaining

exons 2 to 5 might be translated into a truncated version of Oct4 that could retain partial function. Finally, unlike previous study that used maternal-WT zygotic-KO embryos (Stirparo *et al.*, 2021), we used maternal-zygotic-KO embryos (Fig 1C, 1D). The residual levels of maternal Oct4 could have potentially rescued the KO.

Preimplantation embryos undergo dynamic epigenomic reprogramming to regulate the expression of developmental genes. For example, *Ezh2* is required for the establishment and maintenance of global H3K27me3 in preimplantation embryos (Puschendorf *et al*, 2008). Our data indicate that Oct4 and Sox2 act upstream of genes encoding components of epigenetic modifiers, such as *Ezh2* and *Sap30* (Fig 3E, 3F, Supplementary data 2), and thus may be involved this way in regulating the epigenetic status. This suggests a potential amplification of the roles of Oct4 and Sox2 in early development by those downstream epigenetic modifiers.

The core regulatory circuitry of pluripotency exhibits remarkable stability in embryos. Expression of the TFs within this circuitry, such as *Oct4*, *Sox2*, *Nanog*, *Klf4* and *Esrrb*, is partially or fully maintained in the *Oct4*- or *Sox2*-KO ICMs (Le Bin *et al.*, 2014; Wicklow *et al.*, 2014; Wu *et al.*, 2013) (Fig 4C, 4D). The independent and robust expression of these TFs likely ensures the faithful activation of the pluripotency regulatory circuit, safeguarding successful development.

While the cooperative role of Oct4 and Sox2 on the OCT-SOX enhancers in maintaining or inducing the pluripotent network is widely acknowledged, there are still debates regarding the precise mechanisms underlying their cooperation (Biddle *et al.*, 2019; Chen *et al.*, 2008; Han *et al*, 2022; Li *et al.*, 2019; Michael *et al.*, 2020; Velychko *et al*, 2019). In this study, most OCT-SOX peaks are more affected in *Oct4*-KO ICMs than in *Sox2*-KO ICMs, indicating that Oct4 plays a more important role than Sox2 in activating OCT-SOX enhancers. This observation aligns with a previous study showing that the forced expression of Oct4 can rescue pluripotency in *Sox2*-null ESCs (Masui *et al.*, 2007). In addition to their cooperation on OCT-SOX enhancers, recent studies have unveiled independent functions of Oct4 and Sox2 (Friman *et al.*, 2019; Gao *et al.*, 2022). Consistently, we observed that Oct4 regulates a considerable number of peaks and genes independently of Sox2 in the ICMs (Fig 4A, 4C, 4D).

Compared to *Oct4* KO, the loss of *Sox2* resulted in a less severe disruption of the chromatin landscape, transcriptome and phenotype in mouse embryos (Fig 2A-2D, 4A, 4C). We believe that this outcome is unlikely to be attributed to compensation from other members in the Sox family. Sox1, Sox3, Sox15 and Sox18 are promising candidates as they can either rescue the *Sox2*-KO phenotype in mESCs or replace Sox2 to generate iPSCs in reprogramming (Nakagawa *et al*, 2008; Niwa *et al*, 2016). However, our analysis reveals that *Sox1*, *Sox3* and *Sox18* are expressed at extremely low levels in both wildtype and Sox2-KO embryos (Supplementary data 2), suggesting that they are unlikely to be able to fulfill Sox2's role (Deng *et al.*, 2014). Sox15 exhibits indistinguishable expression levels in all three cell types at the blastocyst stage. Furthermore, *Sox15* KO mice displays normal health and fertility (Lee *et al*, 2004). Nonetheless, we cannot completely rule out the possibility of functional compensation by Sox15 in Sox2-KO ICMs.

Our results highlight the crucial role of Oct4 and Sox2 in establishing the transcriptome and chromatin state in the pluripotent EPI. The absence of Sox2 and other factors likely limits the function of Oct4 in morulae. At the blastocyst stage, Oct4 and Sox2 work together to open

the OCT-SOX enhancers and activate pluripotency genes (Fig 6I). However, the upstream factors driving the activation of the core pluripotency regulatory circuitry remain unknown. Further studies are needed to deepen our understanding of the molecular mechanisms governing the embryonic pluripotency program and the early lineage segregation.

Materials and Methods

Mice

To generate ESCs carrying a floxed *Oct4* allele (*Oct4*^{flox}), the construct containing floxed *Oct4* exon 2-5 and a promoter-less FRT-IRES-βgeo-pA cassette was electroporated into germline-competent Acr-EGFP ESCs. Clones were screened for homologous recombination and transiently transfected with an FLP expression vector to remove the FRT cassette. To generate ESCs carrying a mutant *Oct4* allele linked to *mKO2* (*Oct4*^{mKO2}), Acr-EGFP ESCs were targeted with the construct containing a PGK-pacΔtk-P2A-mKO2-pA cassette 3.6 kb upstream of the *Oct4* TSS and a promoter-less FRT-SA-IRES-hph-P2A-Venus-pA cassette in *Oct4* intron 1. To generate ESCs carrying the floxed *Sox2* allele (*Sox2*^{flox}), the construct containing the 5' loxP site 1.9 kb upstream of *Sox2* TSS and a promoter-less FRT-IRES-Neo-pA cassette followed by the 3' loxP site in the 3' UTR of *Sox2* was electroporated into Acr-EGFP ESCs. Clones were screened for homologous recombination and transiently transfected with an FLP expression vector to remove the FRT cassette. To generate ESCs carrying a conditional allele of *Sox2* linked to *EGFP* (*Sox2*^{EGFP}), Acr-EGFP ESCs were targeted with the construct containing the 5' loxP site 1.9 kb upstream of *Sox2* TSS, and a promoter-less FRT-IRES-Neo-pA cassette, the 3' loxP site, and a PGK-EGFP-pA cassette in the 3' UTR of *Sox2*. Chimeric mice were generated by morula aggregation and heterozygous mice were obtained through germline transmission. The Zp3-Cre transgenic mice were used to generate maternal KO alleles. Animal experiments and husbandry were performed according to the German Animal Welfare guidelines and approved by the Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen (State Agency for Nature, Environment and Consumer Protection of North RhineWestphalia).

Embryo collection

All the embryos for RNA-seq, ATAC-seq and microinjection were collected from female mice superovulated with PMSG and hCG. Early and late morula samples were collected at E2.75 and E3.25, respectively. Early and late ICM samples were collected at E2.5 and cultured in KSOM with MEKi (PD0325901, 1 μM) until early and late blastocyst stages, respectively.

ICM isolation

Zona pellucida was removed by brief incubation in prewarmed acidic Tyrode's solution. ICM was obtained by immunosurgery. Briefly, blastocysts were incubated in 20% rabbit anti-mouse whole serum (Sigma-Aldrich) in KSOM at 37°C, 5% CO2 for 15 minutes, followed by three rinses in M2. Afterwards, embryos were incubated in 20% guinea pig complement serum (Sigma-Aldrich) in KSOM at 37°C, 5% CO2 for 15 minutes, followed by three rinses in M2. In the end, ICM was isolated by repetitive blowing in mouth pipette to remove the debris of dead trophectoderm cells.

Immunofluorescence

Immunofluorescence was performed using the following primary antibodies with dilutions: mouse monoclonal anti-Oct4 (sc-5279, Santa Cruz), 1:1000; goat anti-Sox2 (GT15098,

Neuromics), 1:300; rabbit anti-mKO2 (PM051, MBL), 1:1000; rabbit anti-GFP (ab290, Abcam), 1:500; goat anti-Pecam1 (AF3628, R&D biosystems), 1:300; rat monoclonal anti-Nanog (14-5761-80, Thermo Fisher Scientific), 1:100; goat anti-Sox17 (AF1924, R&D Systems), 1:1000 of 0.2 mg/ml.

Microinjection

Zygotes were collected from superovulated B6C3F1 female mice after mating with ICR male mice. Microinjection of mRNA into 1-cell or 2-cell embryos was carried out using the Eppendorf FemtoJet microinjector and Narishige micromanipulators. About 2 pl of mRNAs (10 ng/ul) or plasmids (10 ng/ul) was introduced into the cytoplasm of 1-cell or 2-cell embryos by microinjection. mRNA was synthesized using mMESSAGE mMACHINE T7 Transcription Kit (Thermo Fisher Scientific, AM1344) and poly(A)-tailed using Poly(A) Tailing Kit (Thermo Fisher Scientific, AM1350).

Real-time PCR (RT-PCR)

Real-time PCR was performed using Taq Pro Universal SYBR qPCR Master Mix (Vazyme, Q712-03). The relative expression values were calculated by the ΔΔCt method, normalized to Rplp0. Primer sequences are listed in Table S2.

ATAC-seq and data analysis

ATAC-seq libraries were prepared as previously described with some modifications. 20 to 40 morulae or ICMs were pooled and lysed in the lysis buffer (10 mM Tris·HCl, pH=7.4; 10 mM NaCl; 3 mM MgCl2; 0.15% Igepal CA-630; 0.1% Tween-20) for 10 min on ice. We added 0.1% Tween-20 to the lysis buffer as it greatly reduced reads mapped to mitochondrial DNA. Lyzed embryos were briefly washed, collected in 2.0 µl PBS containing 0.1 mg/ml polyvinyl alcohol (PVA) and transferred to 3.0 µl Tn5 transposome mixture (0.25 µl Tn5 transposome, 2.5 µl tagmentation buffer, Illumina, FC-121-1030; 0.25 µl H_2O). The samples were incubated in 37 °C water bath for 30 min. Tagmentation was stopped by adding 2.0 µl 175 mM EDTA and incubated at 50 °C for 30 min. Excess EDTA was quenched by 2.0 µl 160 mM $MgCl_2$. The libraries were amplified in the following reaction: 9.0 µl Transposed DNA, 10.0 µl NEBNext High-Fidelity 2×PCR Master Mix (New England Biolabs, M0541S), 0.25 µl 100 µM PCR Index 1, 0.25 µl 100 µM PCR Index 2 and 0.5 µl H_2O. The sequences of index 1 and 2 are in Table S3. The PCR program is as the following: 72 °C for 5 min; 98 °C for 30s; 16 thermocycles at 98 °C for 10 s, 63 °C for 30 s and 72 °C for 1 min; followed by 72 °C 5 min. Amplified libraries were purified twice with 1.2 × AMPure XP beads. The sequencing was performed on the NextSeq 500 system with pair-end 75bp.

Sequence reads were trimmed for adapter sequences using SeqPurge and trimmed reads no shorter than 20 bases were aligned to the mm10 mouse reference genome using Bowtie2 with a maximum fragment size of 2000. Duplicated reads were removed using Picard MarkDuplicates (https://broadinstitute.github.io/picard/) and only reads uniquely mapped to the standard chromosomes except chrY and chrM (mitochondrial genome) with mapping quality of at least 30 were used for the following analysis. Reads on the positive and negative strands were shifted +4 bp and −5 bp, respectively. Peak calling was performed using MACS2 with the following parameters: --keep-dup all -g mm --nomodel --shift -50 --extsize 100 -B --SPMR. Peaks that overlap with the blacklisted regions (https://sites.google.com/site/anshulkundaje/projects/blacklists; https://sites.google.com/site/atacseqpublic/atac-seq-analysis-methods/mitochondrialblacklists-1) and satellite repeats (RepeatMasker) were removed.

Significant peak summits (q-value ≤0.001) from biological replicates were merged for each sample group using BEDTools with a maximum distance between two summits to be merged of 200 bp. Only merged peak summits that overlap with summits from at least two replicates were retained. Reads per million (RPM)-normalized pileup signals in the bedGraph format were converted into bigWig files using the UCSC bedGraphToBigWig tool. The bigWig files for the average signal of biological replicates were generated using the UCSC bigWigMerge tool. The heatmaps of RPM-normalized pileup signals around the TSSs of GENCODE (vM23) protein-coding genes were generated using deepTools.

For count-based analysis of transposon insertion events, merged peak regions were generated by combining the adjacent peak summits of all embryo groups and selecting 200-bp regions around the centers of each merged summit. 5' ends of shifted read alignment on both the positive and negative strands were considered as transposon insertion sites. The number of transposon insertions in each merged peak region were counted for each sample using BEDTools and normalized and transformed to $\log_2$ scale using the rlog function of the Bioconductor package DESeq2 and significantly differential peaks were identified with an adjusted p-value cutoff of 0.05. Peaks located within 100 bp from TSSs of GENCODE (vM23) genes were considered as TSS-proximal peaks. The presence of TF binding motifs in merged peaks was investigated using FIMO with the default cutoff (p-value $<10^{-4}$). The enrichment of known TF binding motifs was determined using PScan-ChIP.

RNA-seq and data analysis

Single-embryo RNA-seq was performed as previously described with some modifications (Kurimoto et al. 2007, Nakamura et al. 2015). 2,493 copies of ERCC RNA Spike-In (Ambion) were added to each single-embryo sample. To remove PCR duplicates, we used R2SP-UMI-d(T)24 primer, instead of V1d(T)24 primer, for reverse transcription (RT). To reduce byproducts derived from the RT primer, the poly(A) tailing reaction was performed only for 1 min at 37°C. cDNA was amplified with V3d(T)24 and R2SP primers and Terra PCR Direct Polymerase (Clontech) for 16 cycles. The cDNA was further amplified for four cycles with NH2-V3d(T)24 and NH2-R2SP primers, and purified thrice with 0.6x AMPure XP beads. The 3' end enriched libraries were constructed using KAPA HyperPlus Library Preparation Kit (VWR International, KK8513) with 6 cycles of PCR. All the oligos used in the RT and library amplification are in Table S4. The sequencing was performed on the NextSeq 500 system with single-end 75bp.

UMI sequence of each read was extracted from FASTQ files for Read 2 using UMI-tools. Poly-A sequences were trimmed from the 3' end of Read 1 using Cutadapt (https://cutadapt.readthedocs.org/). The trimmed Read 1 sequences were aligned to the mm10 mouse reference genome using TopHat2 with GENCODE (vM23) transcripts as a transcriptome reference and with default parameters. The reads were also mapped to ERCC reference sequences using Bowtie2 with default parameters. To confirm the genotype of Oct4 and Sox2 KO embryos, we also mapped the reads to the PGK-pacΔtk-P2A-mKO2-hGHpA or PGK-EGFP-hGHpA cassette. The reads mapped to the genome were assigned to GENCODE transcripts using featureCounts. The number of unique transcripts assigned to each GENCODE gene was calculated using TRUmiCount. The transcript counts were normalized and transformed to $\log_2$ scale using the rlog function of the Bioconductor DESeq2 package. Differential expression analysis was performed on genes that were detected in at least half of the samples in at least one stage using DESeq2 and differentially expressed genes were identified with an adjusted p-value <0.05 and $\log_2$ fold change ≥1.

Gene set enrichment analysis was performed to determine whether genes close to the selected ATAC-seq peaks are enriched in genes differentially expressed in *Oct4*- or *Sox2*-KO embryos. The ATAC-seq peaks were annotated to Ensembl protein-coding genes whose TSSs are located within 10 kb of the peak centers using the R package ChIPpeakAnno. FDR was estimated by gene set permutation tests. Heatmaps were generated from normalized enrichment scores (NESs) for gene sets with FDR ≤0.1.

Acknowledgments

We thank Ingrid Gelker, Claudia Ortmeier, David Obridge, Martina Sinn and the animal facility for technical assistance. We are grateful to Dong Han, Rui Fan, Eva Kutejova and Erik Tolen for helpful discussions. We also thank Anika Witten, Christoph Bartenhagen, Carolin Walter and Shixue Gou for their support with the sequencing and RNA-seq data analysis. This work was supported by the Max Planck Society.

Author contributions

K.A. and Y.H. designed the study. Y.H. performed most of the experiments, assembled and interpreted data, and wrote the manuscript. K.A. performed the bioinformatic analysis, interpreted the data and edited the manuscript. G.W. generated the transgenic mice and advised on the project. Z.N. and Y.H. did the microinjection. S.V. and I.B. advised on the project and edited the manuscript. S.H. assisted with the experiments. Q.J. contributed to the bioinformatic analysis. H. Z. edited the manuscript. H.R.S. supervised the project and edited the manuscript.

Conflict of interest

The authors declare no conflicting interests.

References

Adachi K, Kopp W, Wu G, Heising S, Greber B, Stehling M, Arauzo-Bravo MJ, Boerno ST, Timmermann B, Vingron M, Scholer HR (2018) Esrrb Unlocks Silenced Enhancers for Reprogramming to Naive Pluripotency. *Cell Stem Cell*

Adachi K, Nikaido I, Ohta H, Ohtsuka S, Ura H, Kadota M, Wakayama T, Ueda HR, Niwa H (2013) Context-dependent wiring of Sox2 regulatory networks for self-renewal of embryonic and trophoblast stem cells. *Mol Cell* 52: 380-392

Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R (2003) Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* 17: 126-140

Biddle JW, Nguyen M, Gunawardena J (2019) Negative reciprocity, not ordered assembly, underlies the interaction of Sox2 and Oct4 on DNA. *Elife* 8

Boroviak T, Loos R, Lombard P, Okahara J, Behr R, Sasaki E, Nichols J, Smith A, Bertone P (2015) Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. *Developmental cell* 35: 366-382

Buganim Y, Faddah Dina A, Cheng Albert W, Itskovich E, Markoulaki S, Ganz K, Klemm Sandy L, van Oudenaarden A, Jaenisch R (2012) Single-Cell Expression Analyses during Cellular Reprogramming Reveal an Early Stochastic and a Late Hierarchic Phase. *Cell* 150: 1209-1222

Chen J, Zhang Z, Li L, Chen BC, Revyakin A, Hajj B, Legant W, Dahan M, Lionnet T, Betzig E *et al* (2014) Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* 156: 1274-1285

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J *et al* (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133: 1106-1117
Daigneault BW, Rajput S, Smith GW, Ross PJ (2018) Embryonic POU5F1 is Required for Expanded Bovine Blastocyst Formation. *Sci Rep* 8: 7753
Deng Q, Ramskold D, Reinius B, Sandberg R (2014) Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* 343: 193-196
Du Z, Zheng H, Huang B, Ma R, Wu J, Zhang X, He J, Xiang Y, Wang Q, Li Y *et al* (2017) Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* 547: 232-235
Fogarty NME, McCarthy A, Snijders KE, Powell BE, Kubikova N, Blakeley P, Lea R, Elder K, Wamaitha SE, Kim D *et al* (2017) Genome editing reveals a role for OCT4 in human embryogenesis. *Nature* 550: 67-73
Friman ET, Deluz C, Meireles-Filho AC, Govindan S, Gardeux V, Deplancke B, Suter DM (2019) Dynamic regulation of chromatin accessibility by pluripotency transcription factors across the cell cycle. *Elife* 8: e50087
Frum T, Halbisen MA, Wang C, Amiri H, Robson P, Ralston A (2013) Oct4 cell-autonomously promotes primitive endoderm development in the mouse blastocyst. *Developmental cell* 25: 610-622
Frum T, Watts JL, Ralston A (2019) TEAD4, YAP1 and WWTR1 prevent the premature onset of pluripotency prior to the 16-cell stage. *Development (Cambridge, England)* 146
Gao L, Wu K, Liu Z, Yao X, Yuan S, Tao W, Yi L, Yu G, Hou Z, Fan D *et al* (2018) Chromatin Accessibility Landscape in Human Early Embryos and Its Association with Evolution. *Cell* 173: 248-259 e215
Gao M, Veil M, Rosenblatt M, Riesle AJ, Gebhard A, Hass H, Buryanova L, Yampolsky LY, Grüning B, Ulianov SV *et al* (2022) Pluripotency factors determine gene expression repertoire at zygotic genome activation. *Nature Communications* 13: 788
Guo G, Huss M, Tong GQ, Wang C, Li Sun L, Clarke ND, Robson P (2010) Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Developmental cell* 18: 675-685
Han D, Wu G, Chen R, Drexler HCA, MacCarthy CM, Kim K-P, Adachi K, Gerovska D, Mavrommatis L, Bedzhov I *et al* (2022) A balanced Oct4 interactome is crucial for maintaining pluripotency. *Science Advances* 8: eabe4375
Le Bin GC, Munoz-Descalzo S, Kurowski A, Leitch H, Lou X, Mansfield W, Etienne-Dumeau C, Grabole N, Mulas C, Niwa H *et al* (2014) Oct4 is required for lineage priming in the developing inner cell mass of the mouse blastocyst. *Development (Cambridge, England)* 141: 1001-1010
Lee HJ, Göring W, Ochs M, Mühlfeld C, Steding G, Paprotta I, Engel W, Adham IM (2004) Sox15 is required for skeletal muscle regeneration. *Molecular and cellular biology* 24: 8428-8436
Lee MT, Bonneau AR, Takacs CM, Bazzini AA, DiVito KR, Fleming ES, Giraldez AJ (2013) Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* 503: 360-364
Leichsenring M, Maes J, Mossner R, Driever W, Onichtchouk D (2013) Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science* 341: 1005-1009
Li D, Liu J, Yang X, Zhou C, Guo J, Wu C, Qin Y, Guo L, He J, Yu S *et al* (2017) Chromatin Accessibility Dynamics during iPSC Reprogramming. *Cell Stem Cell* 21: 819-833 e816
Li S, Zheng EB, Zhao L, Liu S (2019) Nonreciprocal and Conditional Cooperativity Directs the Pioneer Activity of Pluripotency Transcription Factors. *Cell reports* 28: 2689-2703 e2684
Lu F, Liu Y, Inoue A, Suzuki T, Zhao K, Zhang Y (2016) Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. *Cell* 165: 1375-1388
MacCarthy CM, Malik V, Wu G, Velychko T, Keshet G, Jauch R, Cojocaru V, Schöler HR, Velychko S (2022) Enhancing Sox/Oct cooperativity induces higher-grade developmental reset. *bioRxiv*: 2022.2009. 2023.509242

Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, Guenther MG, Johnston WK, Wernig M, Newman J *et al* (2008) Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 134: 521-533
Masui S, Nakatake Y, Toyooka Y, Shimosato D, Yagi R, Takahashi K, Okochi H, Okuda A, Matoba R, Sharov AA *et al* (2007) Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature cell biology* 9: 625-635
Michael AK, Grand RS, Isbel L, Cavadini S, Kozicka Z, Kempf G, Bunker RD, Schenk AD, Graff-Meyer A, Pathare GR *et al* (2020) Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science* 368: 1460-1465
Minow MAA, Marand AP, Schmitz RJ (2023) Leveraging Single-Cell Populations to Uncover the Genetic Basis of Complex Traits. *Annual review of genetics* 57: null
Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochiduki Y, Takizawa N, Yamanaka S (2008) Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol* 26: 101-106
Ng HH, Surani MA (2011) The transcriptional and signalling networks of pluripotency. *Nature cell biology* 13: 490-496
Nichols J, Silva J, Roode M, Smith A (2009) Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development (Cambridge, England)* 136: 3215-3222
Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Scholer H, Smith A (1998) Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 95: 379-391
Niwa H, Miyazaki J, Smith AG (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature genetics* 24: 372-376
Niwa H, Nakamura A, Urata M, Shirae-Kurabayashi M, Kuraku S, Russell S, Ohtsuka S (2016) The evolutionally-conserved function of group B1 Sox family members confers the unique role of Sox2 in mouse ES cells. *BMC Evol Biol* 16: 173
Palmieri SL, Peter W, Hess H, Schöler HR (1994) Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. *Dev Biol* 166: 259-267
Plachta N, Bollenbach T, Pease S, Fraser SE, Pantazis P (2011) Oct4 kinetics predict cell lineage patterning in the early mammalian embryo. *Nature cell biology* 13: 117-123
Puschendorf M, Terranova R, Boutsma E, Mao X, Isono K, Brykczynska U, Kolb C, Otte AP, Koseki H, Orkin SH *et al* (2008) PRC1 and Suv39h specify parental asymmetry at constitutive heterochromatin in early mouse embryos. *Nature genetics* 40: 411-420
Ralston A, Cox BJ, Nishioka N, Sasaki H, Chea E, Rugg-Gunn P, Guo G, Robson P, Draper JS, Rossant J (2010) Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development (Cambridge, England)* 137: 395-403
Russ AP, Wattler S, Colledge WH, Aparicio SA, Carlton MB, Pearce JJ, Barton SC, Surani MA, Ryan K, Nehls MC *et al* (2000) Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature* 404: 95-99
Stirparo GG, Kurowski A, Yanagida A, Bates LE, Strawbridge SE, Hladkou S, Stuart HT, Boroviak TE, Silva JCR, Nichols J (2021) OCT4 induces embryonic pluripotency via STAT3 signaling and metabolic mechanisms. *Proceedings of the National Academy of Sciences* 118: e2008890118
Strumpf D, Mao CA, Yamanaka Y, Ralston A, Chawengsaksophak K, Beck F, Rossant J (2005) Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development (Cambridge, England)* 132: 2093-2102
Tarkowski AK, Suwińska A, Czołowska R, Ożdżeński W (2010) Individual blastomeres of 16-and 32-cell mouse embryos are able to develop into foetuses and mice. *Developmental biology* 348: 190-198
Velychko S, Adachi K, Kim KP, Hou Y, MacCarthy CM, Wu G, Scholer HR (2019) Excluding Oct4 from Yamanaka Cocktail Unleashes the Developmental Potential of iPSCs. *Cell Stem Cell*

Wang C, Liu X, Gao Y, Yang L, Li C, Liu W, Chen C, Kou X, Zhao Y, Chen J *et al* (2018) Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nature cell biology* 20: 620-631
Wang L, Zhang J, Duan J, Gao X, Zhu W, Lu X, Yang L, Zhang J, Li G, Ci W *et al* (2014) Programming and inheritance of parental DNA methylomes in mammals. *Cell* 157: 979-991
White MD, Angiolini JF, Alvarez YD, Kaur G, Zhao ZW, Mocskos E, Bruno L, Bissiere S, Levi V, Plachta N (2016) Long-Lived Binding of Sox2 to DNA Predicts Cell Fate in the Four-Cell Mouse Embryo. *Cell* 165: 75-87
Whyte WA, Bilodeau S, Orlando DA, Hoke HA, Frampton GM, Foster CT, Cowley SM, Young RA (2012) Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* 482: 221-225
Wicklow E, Blij S, Frum T, Hirate Y, Lang RA, Sasaki H, Ralston A (2014) HIPPO pathway members restrict SOX2 to the inner cell mass where it promotes ICM fates in the mouse blastocyst. *PLoS genetics* 10: e1004618
Wu G, Han D, Gong Y, Sebastiano V, Gentile L, Singhal N, Adachi K, Fischedick G, Ortmeier C, Sinn M *et al* (2013) Establishment of totipotency does not depend on Oct4A. *Nature cell biology* 15: 1089-1097
Wu J, Huang B, Chen H, Yin Q, Liu Y, Xiang Y, Zhang B, Liu B, Wang Q, Xia W *et al* (2016) The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* 534: 652-657
Ying Q-L, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, Smith A (2008) The ground state of embryonic stem cell self-renewal. *Nature* 453: 519-523
Young RA (2011) Control of the embryonic stem cell state. *Cell* 144: 940-954
Zhang B, Zheng H, Huang B, Li W, Xiang Y, Peng X, Ming J, Wu X, Zhang Y, Xu Q *et al* (2016) Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* 537: 553-557
Zhang Y, Xiang Y, Yin Q, Du Z, Peng X, Wang Q, Fidalgo M, Xia W, Li Y, Zhao ZA *et al* (2018) Dynamic epigenomic landscapes during early lineage specification in mouse embryos. *Nature genetics* 50: 96-105

Figure legends

Fig 1 Generation of maternal-zygotic KO embryos. A. Schemes of mKO2-labeled *Oct4* KO (*Oct4*^{mKO2}) and *Oct4* ^{flox} alleles. **B**. Schemes of EGFP-labeled *Sox2* KO (*Sox2*^{EGFP}) and *Sox2* ^{flox} alleles. **C**. Mating strategy for *Oct4* Ctrl and KO embryos. **D**. Mating strategy for *Sox2* Ctrl and KO embryos. **E**. Immunostaining with embryos separated based on the mKO2 fluorescence. **F**. Immunostaining with embryos separated based on the EGFP fluorescence. N, the number of embryos. Scalebar, 20 μm.

Fig 2 Oct4 and Sox2 regulate chromatin landscape and transcriptome in the ICM, but not in the morula. A. PCA plot of ATAC-seq samples. **B**. k-means clustering of the ATAC-seq peaks across all the samples. The heatmap is sorted by clusters, motifs and TSS. **C**. Number of differentially accessible ATAC-seq peaks in KO *vs* Ctrl samples. **D**. Number of differentially expressed genes in KO *vs* Ctrl samples.

Fig 3 Altered chromatin accessibility results in transcriptional changes. A&B. GSEA plots show the transcriptional activity of the genes within 10 kb of the significantly increased and decreased ATAC-seq peak in *Oct4*-KO and *Sox2*-KO late ICMs. **C&D**. Heatmap of GSEA enrichment in all the samples. The number in the parenthesis represents the number of genes whose TSS was located within 10 kb of the significantly changed peaks in each module. **E.** Examples of down- and up-regulated genes. The underlined numbers represent the adjusted p values. **F**. The ATAC-seq profiles surrounding the genes in E. Boxes mark the differentially accessible peaks and red boxes mark those with the OCT-SOX motif. mm, mismatch.

Fig 4. Oct4 and Sox2 activate EPI-specific genes and suppress TE-specific genes in ICM. A. A scatter plot shows the $\log_2$ fold change of ATAC-seq signals in the *Oct4*-KO and *Sox2*-KO late ICM. Colored dots represent the significantly changed peaks. **B**. GREAT ontology enrichment analysis of the significantly changed peaks shared in the *Oct4*-KO and *Sox2*-KO late ICM. Red terms are related to the pluripotency and preimplantation embryonic development, and green ones are related to the development of embryonic and extraembryonic lineages in the post-implantation embryos. **C**. A scatter plot shows the $\log_2$ fold changes of RNA-seq signals in the *Oct4*-KO and *Sox2*-KO late ICM. **D**. Violin plots of RNA-seq rlog values for *Nanog*, *Esrrb* and *Klf4* in the embryos. The underlined numbers represent the adjusted p values. **E**. Immunostaining of Pecam1 in the late blastocysts (E2.5+2 days). Scalebar, 50 μm. **F**. The accessibility of putative enhancers around *Pecam1* in ICMs. Boxes mark the decreased peaks and red box marks the peak with the OCT-SOX motif.

Fig 5 Oct4 and Sox2 activate OCT-SOX enhancers cooperatively and independently in ICM. A. Motif enrichment analysis of significantly changed ATAC-seq peaks in the *Oct4*-KO and *Sox2*-KO ICMs. **B**. Occurrence of the motifs in the ranked peaks. The decreased enhancers were ranked by the fold reduction. The cumulative percentages of peaks containing at least one sequence for a given motif are plotted against ranks of peaks. **C**. Scatter plots show the $\log_2$ fold change of 8,993 OCT-SOX peaks in the *Oct4*-KO and *Sox2*-KO early and late ICMs. **D**. The rlog values of the OCT-SOX peaks in C. **E**. The canonical OCT-SOX motif. Boxes mark the highly conserved bases in the Oct4 POU-specific domain (GCAT) and Sox2-binding (ACAA) site. **F**. Violin plots show the rlog values of the four groups of OCT-SOX peaks classified based on the presence of mismatches (mm) to the conserved bases. **G**. The colocalization of accessible chromatin regions in embryo and TF binding in ESCs around the OCT-SOXmm enhancers of *Esrrb* and *Klf4*. The red boxes mark the OCT-SOXmm enhancers with their sequences are shown above. The numbers represent the $\log_2$ fold change of ATAC-seq signals in the KO *vs* Ctrl ICMs.

Fig 6 Sox2 acts as a rate-limiting factor for the activation of the EPI-lineage genes. **A&B**. Alluvial plots show the dynamic changes of ATAC-seq (A) and RNA-seq (B) signals from morula to late ICM in the Ctrl embryos. Green, grey and red lines represent the decreased, unchanged and increased peaks/genes, respectively. **C**. Expression of *Sox2*, *Oct4*, *Nanog*, *Esrrb* and *Klf4* in morulae and ICMs. **D**. The bar chart illustrates the enrichment results obtained through the GSEA method using the Wikipathway functional database in WebGestalt. The $\log_2$ fold changes of the 2,462 upregulated genes (Fig 6B) in Sox2 KO *vs* Ctrl early ICM were used in this analysis. **E**. Enrichment plot of the term PluriNetWork in Fig 6D. **F**. Scheme for the microinjection of mRNA into zygote cytoplasm. **G**. Overexpression of *Sox2* was confirmed by immunostaining in morulae at E0.5 + 2 days after the microinjection. Scale bar, 50 μm. **H**. A heatmap shows the mean-centered ΔCt values of the EPI and TE makers in the *Sox2*- and *mCherry*-overexpressed morula. **I**. Model of Sox2 as a rate-limiting factor for the activation of pluripotency-related genes in the EPI.

Supplementary figure legends

Fig S1 *Oct4* or *Sox2* Ctrl and KO embryos under florescent microscopy (related to Fig 1). A. The mKO2-labeled *Oct4*-KO embryos were identified under fluorescent microscopy at E2.5+1 day. **B**. The EGFP-labeled *Sox2*-KO embryos were identified under florescent microscopy at E2.5+2 days.

Fig S2 MEKi specified the ICM to EPI fate (related to Fig 2). A. The scheme of MEKi (PD0325901, 1 μM) treatment. **B.** Immunostaining of Sox17 and Nanog in the late blastocysts treated with or without MEKi. **C.** The number of Sox17+ or Nanog+ cells in each blastocyst.

Fig S3 Quality check of the low-input ATAC-seq and RNA-seq (related to Fig 2). A. The location of ATAC-seq peaks relative to TSSs. The number in each bar represents the number of peaks in each group. **B.** ChIP-seq profiles of Oct4, Sox2, Nanog (Marson *et al.*, 2008) and H3K27ac (Whyte *et al.*, 2012) in ESCs over the ATAC-seq peaks. **C.** PCA plot of RNA-seq samples with the indicated genotypes.

Fig S4 GREAT ontology enrichment analysis of ATAC-seq peaks decreased in the *Oct4*-KO and unchanged in the *Sox2*-KO late ICMs (related to Fig 4). Terms highlighted in red are related to the pluripotency and preimplantation embryonic development. Example genes associated with the terms are listed.

Fig S5 Motif enrichment analysis of significantly changed ATAC-seq peaks that are distal to TSSs (related to Fig 5). The values are $-\log_{10}$(p-value).

Fig S6 MA plots show the $\log_2$ fold change of the 8,993 OCT-SOX peaks in Oct4-KO and Sox2-KO ICMs (related to Fig 5). Black and red dots represent the unchanged and significantly changed peaks, respectively. Grey lines are the regression curves of the 8,993 OCT-SOX peaks.

Fig S7 The dynamics of chromatin accessibility and transcriptome from morula to ICM (related to Fig 6). A GREAT analysis was conducted on the significantly increased ATAC-seq peaks in *Ctrl* early ICM *vs Ctrl* morula. A log2 fold change cutoff of >2 was applied due to the limits of GREAT analysis with fewer than 10,000 peaks. **B**. Motif enrichment analysis of the top increased ATAC-seq peaks from morula to early ICM. **C**. Approximately 50% of the diminished peaks (left) and genes (right) in Oct4-KO and Sox2-KO early ICMs were augmented from the morula to the ICM stage.

Table S1 Coordinates for marked enhancers in Fig 3A

Table S2 Primer sequences for RT-PCR
Table S3 oligos for ATAC-seq
Table S4 oligos for RNA -seq
Supplementary data 1 DESeq2 of ATAC-seq
Supplementary data 2 DESeq2 of RNA-seq

Figure 1

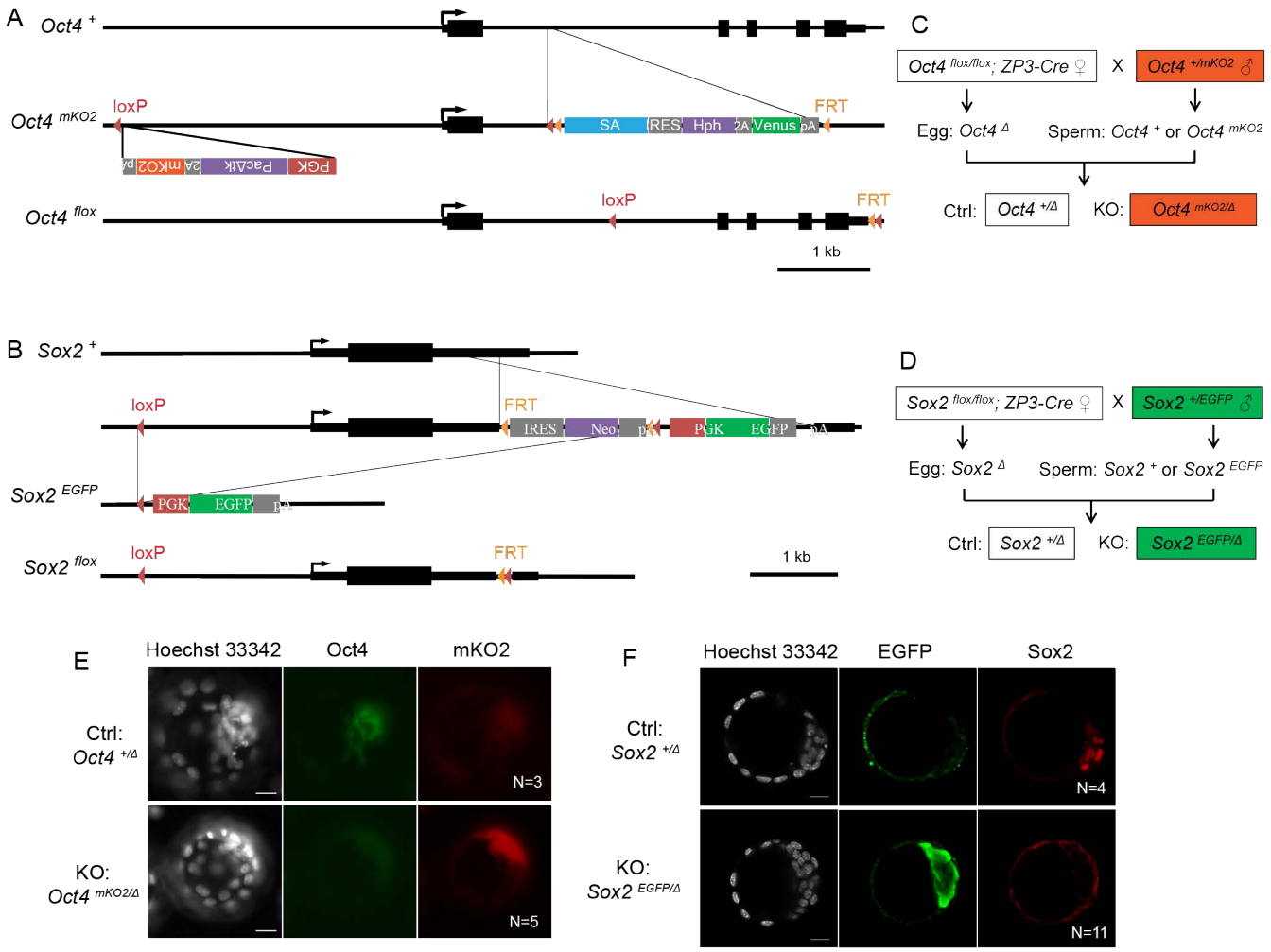
A
Oct4 +
Oct4 mKO2
loxP
FRT
SA
IRES
Hph
2A
Venus
pA
PGK
PacAtk
2A
mKO2
pA
Oct4 flox
loxP
FRT
1 kb
C
Oct4 flox/flox; ZP3-Cre ♀
X
Oct4 +/mKO2 ♂
Egg: Oct4 Δ
Sperm: Oct4 + or Oct4 mKO2
Ctrl: Oct4 +/Δ
KO: Oct4 mKO2/Δ
B
Sox2 +
loxP
FRT
IRES
Neo
pA
PGK
EGFP
pA
Sox2 EGFP
PGK
EGFP
pA
Sox2 flox
loxP
FRT
1 kb
D
Sox2 flox/flox; ZP3-Cre ♀
X
Sox2 +/EGFP ♂
Egg: Sox2 Δ
Sperm: Sox2 + or Sox2 EGFP
Ctrl: Sox2 +/Δ
KO: Sox2 EGFP/Δ
E
Hoechst 33342
Oct4
mKO2
Ctrl:
Oct4 +/Δ
N=3
KO:
Oct4 mKO2/Δ
N=5
F
Hoechst 33342
EGFP
Sox2
Ctrl:
Sox2 +/Δ
N=4
KO:
Sox2 EGFP/Δ
N=11

Figure 1 Generation of maternal-zygotic KO embryos. A. Schemes of mKO2-labeled *Oct4* KO (*Oct4*^{mKO2}) and *Oct4* ^{flox} alleles. **B**. Schemes of EGFP-labeled *Sox2* KO (*Sox2*^{EGFP}) and *Sox2* ^{flox} alleles. **C**. Mating strategy for *Oct4* Ctrl and KO embryos. **D**. Mating strategy for *Sox2* Ctrl and KO embryos. **E**. Immunostaining with embryos separated based on the mKO2 fluorescence. **F**. Immunostaining with embryos separated based on the EGFP fluorescence. N, the number of embryos. Scale bar, 20 μm.

Figure 2

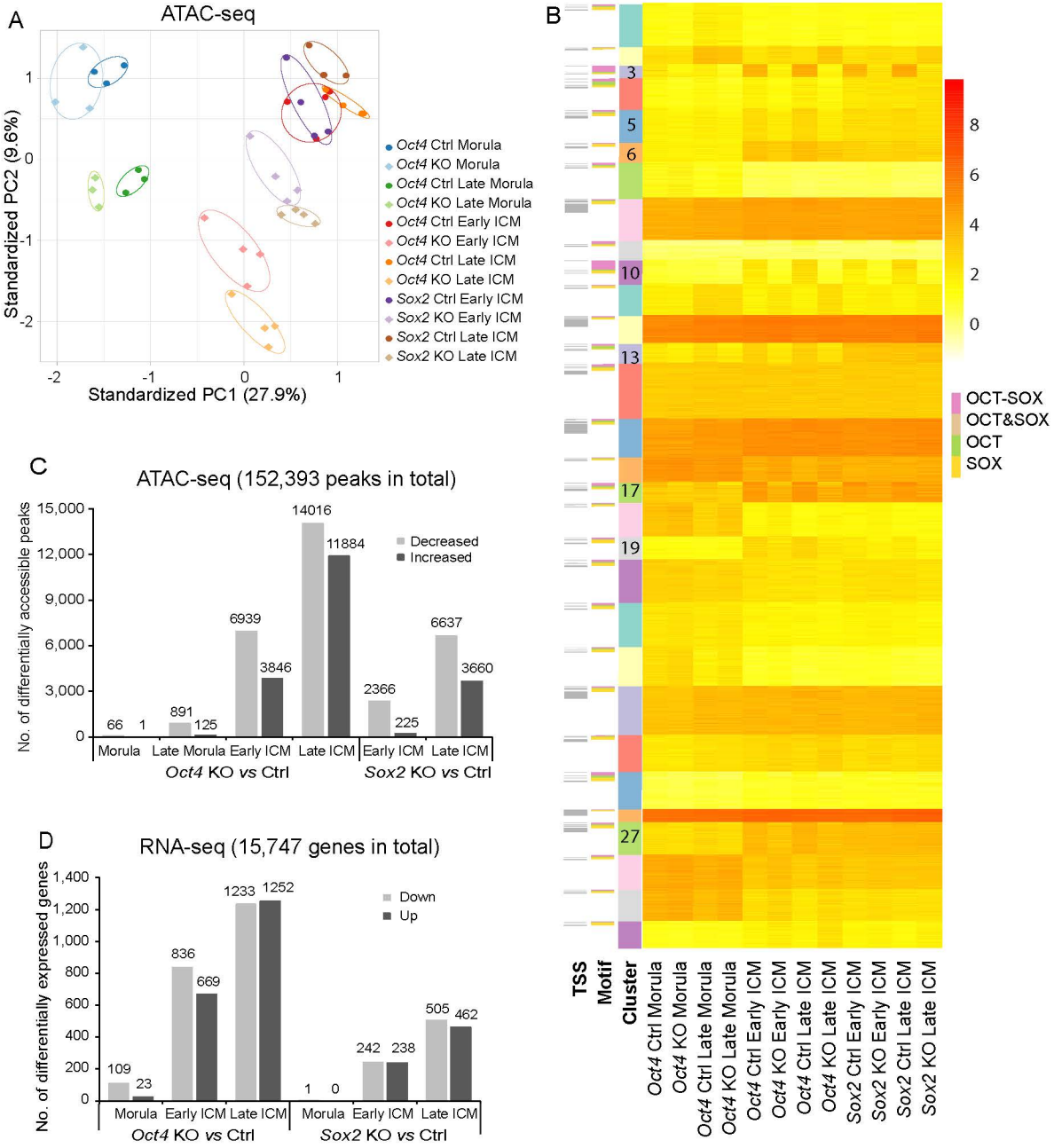
A
ATAC-seq
Standardized PC2 (9.6%)
Standardized PC1 (27.9%)
Oct4 Ctrl Morula
Oct4 KO Morula
Oct4 Ctrl Late Morula
Oct4 KO Late Morula
Oct4 Ctrl Early ICM
Oct4 KO Early ICM
Oct4 Ctrl Late ICM
Oct4 KO Late ICM
Sox2 Ctrl Early ICM
Sox2 KO Early ICM
Sox2 Ctrl Late ICM
Sox2 KO Late ICM
B
TSS
Motif
Cluster
OCT-SOX
OCT&SOX
OCT
SOX
C
ATAC-seq (152,393 peaks in total)
No. of differentially accessible peaks
Decreased
Increased
Morula
Late Morula
Early ICM
Late ICM
Oct4 KO vs Ctrl
Sox2 KO vs Ctrl
D
RNA-seq (15,747 genes in total)
No. of differentially expressed genes
Down
Up

Fig 2 Oct4 and Sox2 regulate chromatin landscape and transcriptome in the ICM, but not in the morula. A. PCA plot of ATAC-seq samples. **B**. k-means clustering of the ATAC-seq peaks across all the samples. The heatmap is sorted by clusters, motifs and TSS. **C**. Number of differentially accessible ATAC-seq peaks in KO *vs* Ctrl samples. **D**. Number of differentially expressed genes in KO *vs* Ctrl samples.

Figure 3

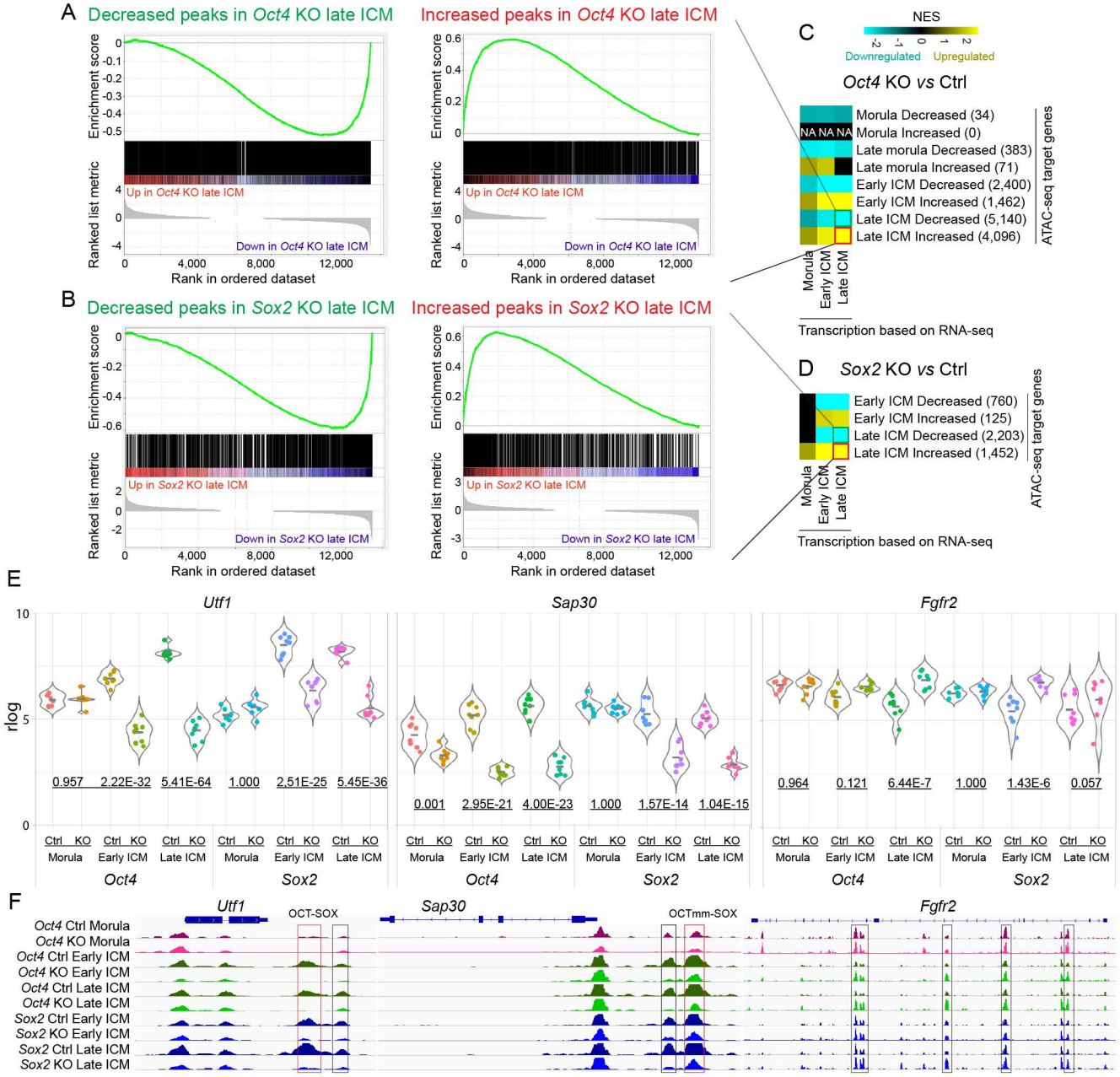
A
Decreased peaks in Oct4 KO late ICM
Increased peaks in Oct4 KO late ICM
Enrichment score
Ranked list metric
Up in Oct4 KO late ICM
Down in Oct4 KO late ICM
Rank in ordered dataset
B
Decreased peaks in Sox2 KO late ICM
Increased peaks in Sox2 KO late ICM
Up in Sox2 KO late ICM
Down in Sox2 KO late ICM
C
NES
Downregulated
Upregulated
Oct4 KO vs Ctrl
Morula Decreased (34)
Morula Increased (0)
Late morula Decreased (383)
Late morula Increased (71)
Early ICM Decreased (2,400)
Early ICM Increased (1,462)
Late ICM Decreased (5,140)
Late ICM Increased (4,096)
ATAC-seq target genes
Morula
Early ICM
Late ICM
Transcription based on RNA-seq
D
Sox2 KO vs Ctrl
Early ICM Decreased (760)
Early ICM Increased (125)
Late ICM Decreased (2,203)
Late ICM Increased (1,452)
ATAC-seq target genes
Transcription based on RNA-seq
E
Utf1
Sap30
Fgfr2
rlog
0.957 2.22E-32 5.41E-64 1.000 2.51E-25 5.45E-36
0.001 2.95E-21 4.00E-23 1.000 1.57E-14 1.04E-15
0.964 0.121 6.44E-7 1.000 1.43E-6 0.057
Ctrl KO
Morula
Early ICM
Late ICM
Oct4
Sox2
F
OCT-SOX
OCTmm-SOX
Oct4 Ctrl Morula
Oct4 KO Morula
Oct4 Ctrl Early ICM
Oct4 KO Early ICM
Oct4 Ctrl Late ICM
Oct4 KO Late ICM
Sox2 Ctrl Early ICM
Sox2 KO Early ICM
Sox2 Ctrl Late ICM
Sox2 KO Late ICM

Figure 3 Altered chromatin accessibility results in transcriptional changes. A&B. GSEA plots show the transcriptional activity of the genes within 10 kb of the significantly increased and decreased ATAC-seq peak in *Oct4*-KO and *Sox2*-KO late ICMs. **C&D**. Heatmap of GSEA enrichment in all the samples. The number in the parenthesis represents the number of genes whose TSS was located within 10 kb of the significantly changed peaks in each module. **E.** Examples of down- and up-regulated genes. The underlined numbers represent the adjusted p values. **F**. The ATAC-seq profiles surrounding the genes in E. Boxes mark the differentially accessible peaks and red boxes mark those with the OCT-SOX motif. mm, mismatch.

Figure 4

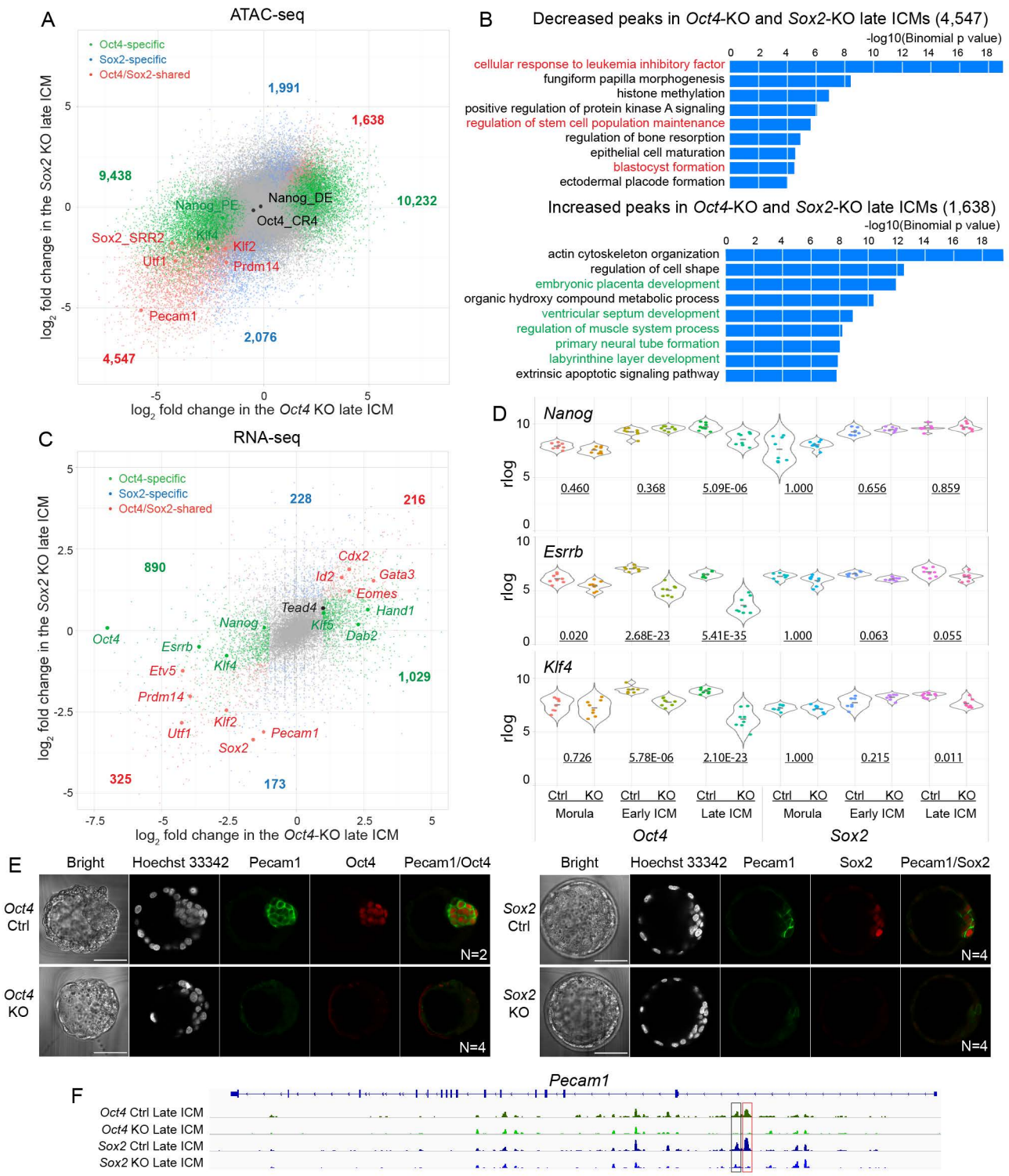
A
ATAC-seq
Oct4-specific
Sox2-specific
Oct4/Sox2-shared
1,991
1,638
9,438
10,232
Nanog_PE
Nanog_DE
Oct4_CR4
Sox2_SRR2
Klf4
Klf2
Utf1
Prdm14
Pecam1
2,076
4,547
log2 fold change in the Sox2 KO late ICM
log2 fold change in the Oct4 KO late ICM
B
Decreased peaks in Oct4-KO and Sox2-KO late ICMs (4,547)
-log10(Binomial p value)
cellular response to leukemia inhibitory factor
fungiform papilla morphogenesis
histone methylation
positive regulation of protein kinase A signaling
regulation of stem cell population maintenance
regulation of bone resorption
epithelial cell maturation
blastocyst formation
ectodermal placode formation
Increased peaks in Oct4-KO and Sox2-KO late ICMs (1,638)
actin cytoskeleton organization
regulation of cell shape
embryonic placenta development
organic hydroxy compound metabolic process
ventricular septum development
regulation of muscle system process
primary neural tube formation
labyrinthine layer development
extrinsic apoptotic signaling pathway
C
RNA-seq
228
216
890
Cdx2
Id2
Gata3
Eomes
Tead4
Hand1
Nanog
Klf5
Dab2
Oct4
Esrrb
Etv5
Klf4
1,029
Prdm14
Klf2
Utf1
Pecam1
Sox2
325
173
log2 fold change in the Oct4-KO late ICM
D
Nanog
Esrrb
Klf4
rlog
0.460
0.368
5.09E-06
1.000
0.656
0.859
0.020
2.68E-23
5.41E-35
1.000
0.063
0.055
0.726
5.78E-06
2.10E-23
1.000
0.215
0.011
Ctrl KO
Morula
Early ICM
Late ICM
Oct4
Sox2
E
Bright
Hoechst 33342
Pecam1
Oct4
Pecam1/Oct4
Oct4 Ctrl
Oct4 KO
N=2
N=4
Sox2
Pecam1/Sox2
Sox2 Ctrl
Sox2 KO
F
Pecam1
Oct4 Ctrl Late ICM
Oct4 KO Late ICM
Sox2 Ctrl Late ICM
Sox2 KO Late ICM

Figure 4. Oct4 and Sox2 activate EPI-specific genes and suppress TE-specific genes in ICM. A. A scatter plot shows the $\log_2$ fold change of ATAC-seq signals in the *Oct4*-KO and *Sox2*-KO late ICM. Colored dots represent the significantly changed peaks. **B**. GREAT ontology enrichment analysis of the significantly changed peaks shared in the *Oct4*-KO and *Sox2*-KO late ICM. Red terms are related to the pluripotency and preimplantation embryonic development, and green ones are related to the development of embryonic and extraembryonic lineages in the post-implantation embryos. **C**. A scatter plot shows the $\log_2$ fold changes of RNA-seq signals in the *Oct4*-KO and *Sox2*-KO late ICM. **D**. Violin plots of RNA-seq rlog values for *Nanog*, *Esrrb* and *Klf4* in the embryos. The underlined numbers represent the adjusted p values. **E**. Immunostaining of Pecam1 in the late blastocysts (E2.5+2 days). Scalebar, 50 μm. **F**. The accessibility of putative enhancers around *Pecam1* in ICMs. Boxes mark the decreased peaks and red box marks the peak with the OCT-SOX motif.

Figure 5

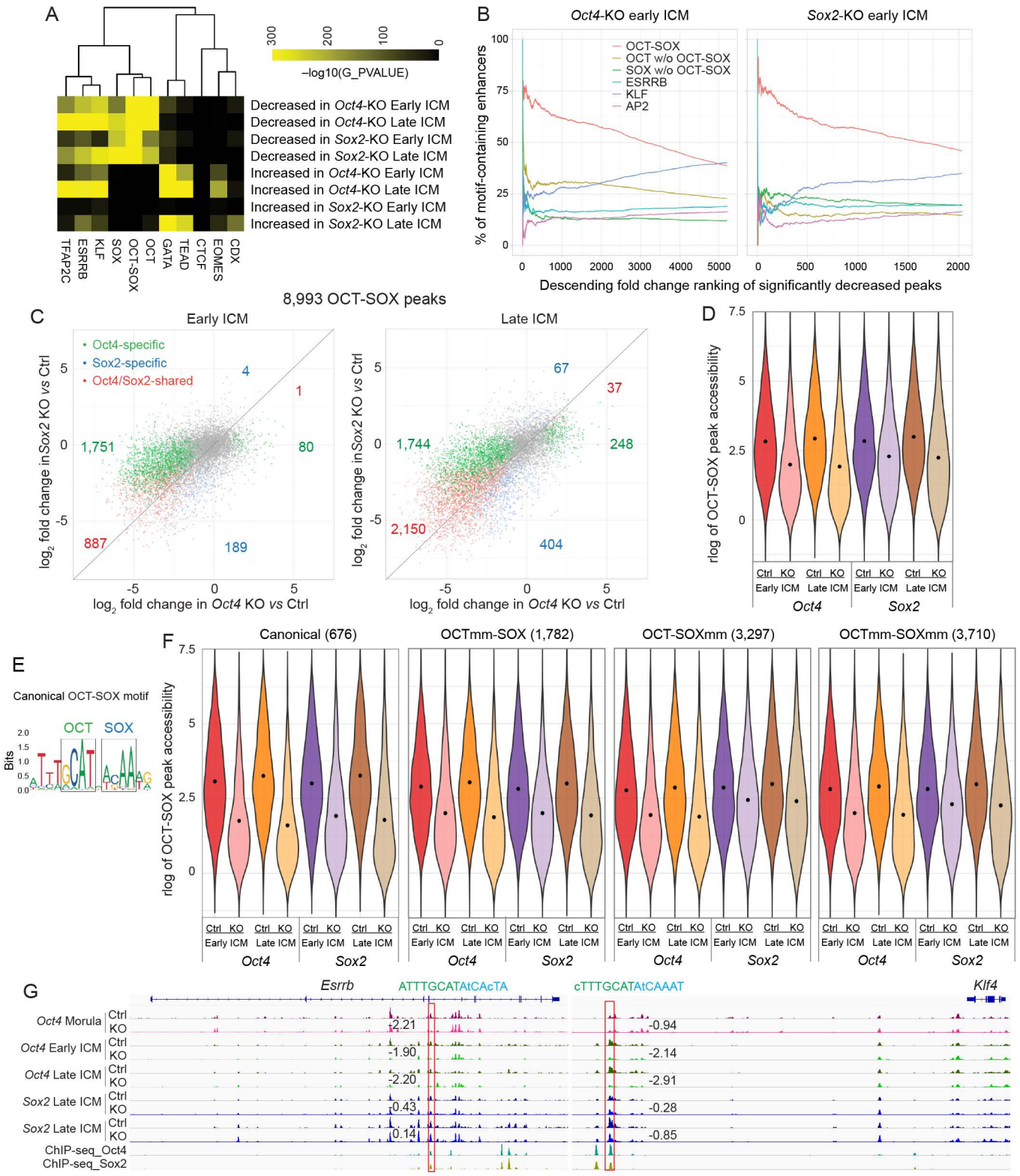
A
−log10(G_PVALUE)
Decreased in Oct4-KO Early ICM
Decreased in Oct4-KO Late ICM
Decreased in Sox2-KO Early ICM
Decreased in Sox2-KO Late ICM
Increased in Oct4-KO Early ICM
Increased in Oct4-KO Late ICM
Increased in Sox2-KO Early ICM
Increased in Sox2-KO Late ICM
TFAP2C
ESRRB
KLF
SOX
OCT-SOX
OCT
GATA
TEAD
CTCF
EOMES
CDX
B
Oct4-KO early ICM
Sox2-KO early ICM
OCT-SOX
OCT w/o OCT-SOX
SOX w/o OCT-SOX
ESRRB
KLF
AP2
% of motif-containing enhancers
Descending fold change ranking of significantly decreased peaks
8,993 OCT-SOX peaks
C
Early ICM
Late ICM
Oct4-specific
Sox2-specific
Oct4/Sox2-shared
log2 fold change in Sox2 KO vs Ctrl
log2 fold change in Oct4 KO vs Ctrl
D
rlog of OCT-SOX peak accessibility
Ctrl KO Ctrl KO Ctrl KO Ctrl KO
Early ICM Late ICM Early ICM Late ICM
Oct4
Sox2
E
Canonical OCT-SOX motif
OCT
SOX
Bits
F
Canonical (676)
OCTmm-SOX (1,782)
OCT-SOXmm (3,297)
OCTmm-SOXmm (3,710)
G
Esrrb
ATTTGCATAtCAcTA
cTTTGCATAtCAAAT
Klf4
Oct4 Morula
Oct4 Early ICM
Oct4 Late ICM
Sox2 Late ICM
Sox2 Late ICM
Ctrl
KO
-2.21
-0.94
-1.90
-2.14
-2.20
-2.91
-0.43
-0.28
-0.14
-0.85
ChIP-seq_Oct4
ChIP-seq_Sox2

Figure 5 Oct4 and Sox2 activate OCT-SOX enhancers cooperatively and independently in ICM. A. Motif enrichment analysis of significantly changed ATAC-seq peaks in the *Oct4*-KO and *Sox2*-KO ICMs. **B**. Occurrence of the motifs in the ranked peaks. The decreased enhancers were ranked by the fold reduction. The cumulative percentages of peaks containing at least one sequence for a given motif are plotted against ranks of peaks. **C**. Scatter plots show the $\log_2$ fold change of 8,993 OCT-SOX peaks in the *Oct4*-KO and *Sox2*-KO early and late ICMs. **D**. The rlog values of the OCT-SOX peaks in C. **E**. The canonical OCT-SOX motif. Boxes mark the highly conserved bases in the Oct4 POU-specific domain (GCAT) and Sox2-binding (ACAA) site. **F**. Violin plots show the rlog values of the four groups of OCT-SOX peaks classified based on the presence of mismatches (mm) to the conserved bases. **G**. The colocalization of accessible chromatin regions in embryo and TF binding in ESCs around the OCT-SOXmm enhancers of *Esrrb* and *Klf4*. The red boxes mark the OCT-SOXmm enhancers with their sequences are shown above. The numbers represent the $\log_2$ fold change of ATAC-seq signals in the KO *vs* Ctrl ICMs.

Figure 6

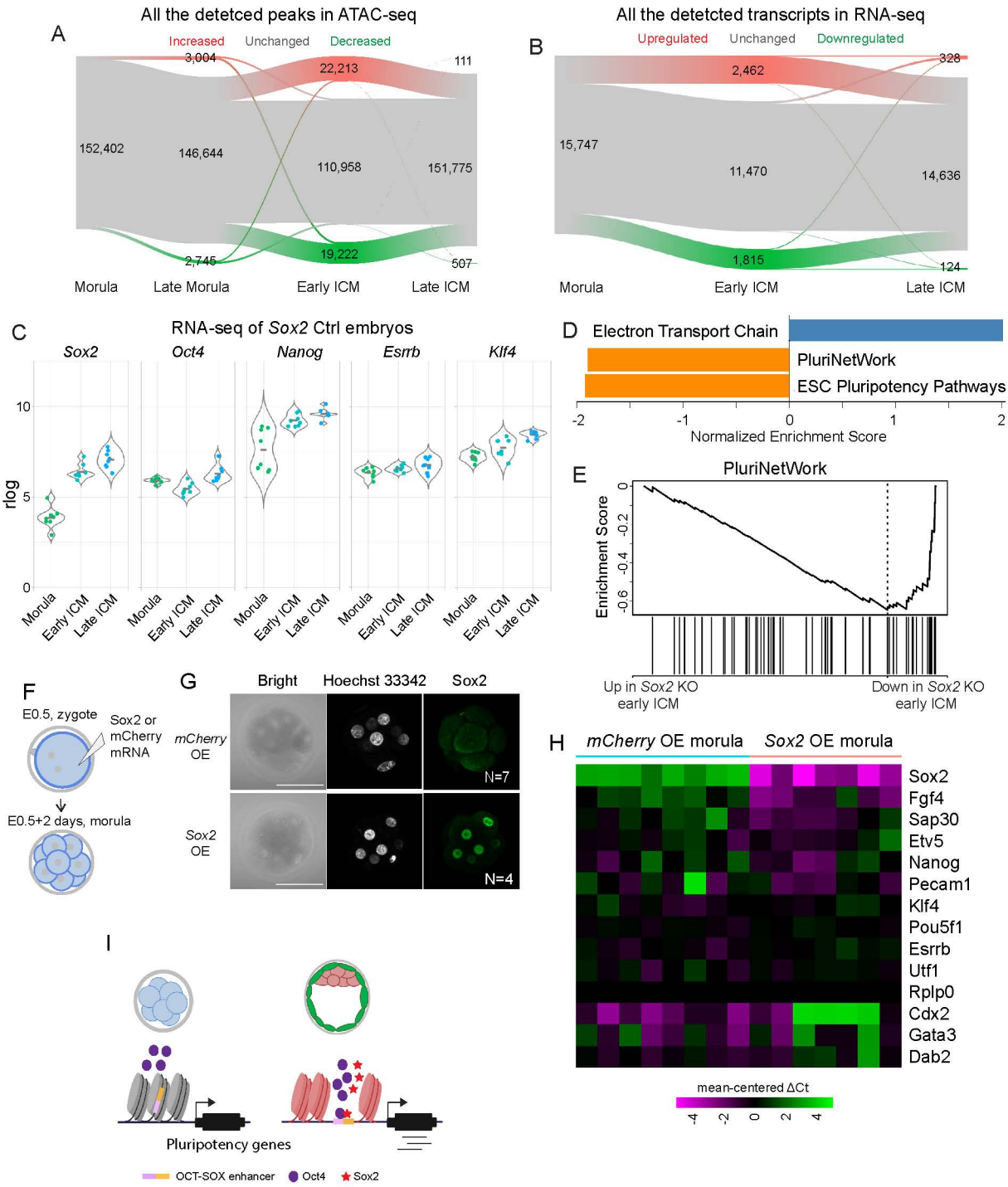
A
All the detetced peaks in ATAC-seq
Increased
Unchanged
Decreased
3,004
22,213
111
152,402
146,644
110,958
151,775
2,745
19,222
507
Morula
Late Morula
Early ICM
Late ICM
B
All the detetcted transcripts in RNA-seq
Upregulated
Unchanged
Downregulated
2,462
328
15,747
11,470
14,636
1,815
124
Morula
Early ICM
Late ICM
C
RNA-seq of Sox2 Ctrl embryos
Sox2
Oct4
Nanog
Esrrb
Klf4
rlog
D
Electron Transport Chain
PluriNetWork
ESC Pluripotency Pathways
Normalized Enrichment Score
E
PluriNetWork
Enrichment Score
Up in Sox2 KO early ICM
Down in Sox2 KO early ICM
F
E0.5, zygote
Sox2 or mCherry mRNA
E0.5+2 days, morula
G
Bright
Hoechst 33342
Sox2
mCherry OE
Sox2 OE
N=7
N=4
H
mCherry OE morula
Sox2 OE morula
Sox2
Fgf4
Sap30
Etv5
Nanog
Pecam1
Klf4
Pou5f1
Esrrb
Utf1
Rplp0
Cdx2
Gata3
Dab2
mean-centered ΔCt
I
Pluripotency genes
OCT-SOX enhancer
Oct4
Sox2

Figure 6 Sox2 acts as a rate-limiting factor for the activation of the EPI-lineage genes. **A&B**. Alluvial plots show the dynamic changes of ATAC-seq (A) and RNA-seq (B) signals from morula to late ICM in the Ctrl embryos. Green, grey and red lines represent the decreased, unchanged and increased peaks/genes, respectively. **C**. Expression of *Sox2*, *Oct4*, *Nanog*, *Esrrb* and *Klf4* in morulae and ICMs. **D**. The bar chart illustrates the enrichment results obtained through the GSEA method using the Wikipathway functional database in WebGestalt. The $\log_2$ fold changes of the 2,462 upregulated genes (Figure 6B) in Sox2 KO *vs* Ctrl early ICM were used in this analysis. **E**. Enrichment plot of the term PluriNetWork in Figure 6D. **F**. Scheme for the microinjection of mRNA into zygote cytoplasm. **G**. Overexpression of *Sox2* was confirmed by immunostaining in morulae at E0.5 + 2 days after the microinjection. Scale bar, 50 μm. **H**. A heatmap shows the mean-centered ΔCt values of the EPI and TE makers in the *Sox2*- and *mCherry*-overexpressed morula. **I**. Model of Sox2 as a rate-limiting factor for the activation of pluripotency-related genes in mouse embryos.