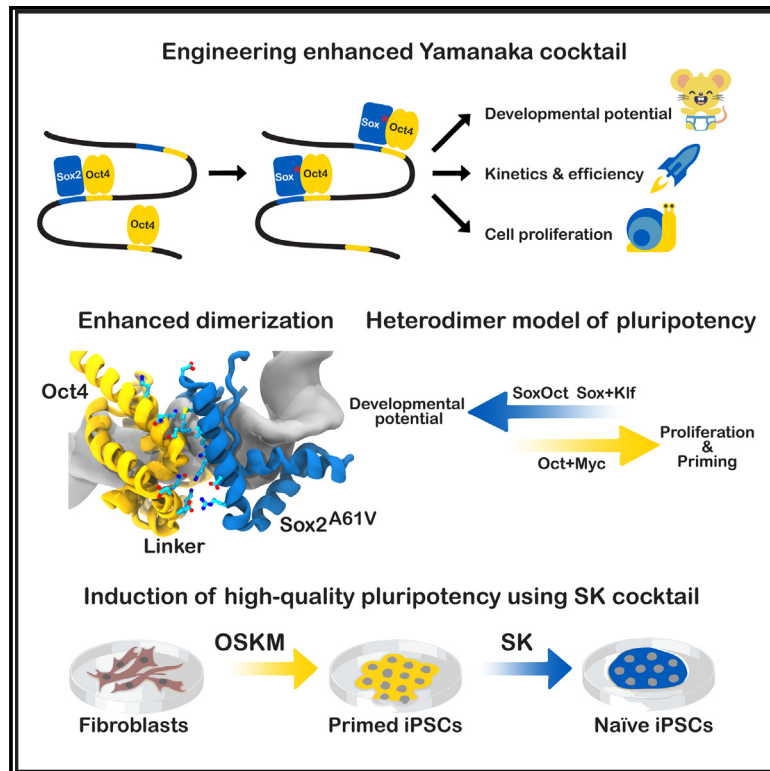# Highly cooperative chimeric super-SOX induces naive pluripotency across species

Graphical abstract

Authors

Caitlin M. MacCarthy, Guangming Wu, Vikas Malik, ..., Vlad Cojocaru, Hans R. Schöler, Sergiy Velychko

Correspondence

h.schoeler@mpi-muenster.mpg.de (H.R.S.), sergiy_velychko@hms.harvard.edu (S.V.)

In brief

Certain structural elements of Sox17 could enhance Sox2's ability to generate iPSCs by stabilizing Sox2/Oct4 dimerization on regulatory DNA elements that control pluripotency. This study highlights an engineered super-reprogramming factor Sox2-17 and reveals the key mechanism driving complete developmental reset.

Highlights

- Yamanaka cocktail upgraded with Sox2-17 enhanced iPSC generation in 5 species

- $Sox2^{A61 \to V}$ swap at Sox2/Oct4 interface improves the quality of OSKM iPSCs

- Sox2/Oct4 dimer is the master regulator of high-quality naive pluripotency

- Naive reset using episomal SK boosts the developmental potential of PSCs across species

Cell Stem Cell

Article

Highly cooperative chimeric super-SOX induces naive pluripotency across species

Caitlin M. MacCarthy,[1] Guangming Wu,[1,2,3] Vikas Malik,[4] Yotam Menuchin-Lasowski,[1] Taras Velychko,[1,5] Gal Keshet,[6] Rui Fan,[1] Ivan Bedzhov,[1] George M. Church,[7,8] Ralf Jauch,[9,10] Vlad Cojocaru,[1,11,12] Hans R. Schöler,[1,*] and Sergiy Velychko[1,7,8,13,*]

[1]Max Planck Institute for Molecular Biomedicine, Münster, Germany
[2]International Bio Island, Guangzhou, China
[3]MingCeler Biotech, Guangzhou, China
[4]Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA
[5]Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany
[6]Hebrew University of Jerusalem, Jerusalem, Israel
[7]Department of Genetics, Harvard Medical School, Boston, MA, USA
[8]Wyss Institute, Harvard University, Boston, MA, USA
[9]School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China
[10]Centre for Translational Stem Cell Biology, Hong Kong SAR, China
[11]University of Utrecht, Utrecht, the Netherlands
[12]STAR-UBB Institute, Babeş-Bolyai University, Cluj-Napoca, Romania
[13]Lead contact
*Correspondence: h.schoeler@mpi-muenster.mpg.de (H.R.S.), sergiy_velychko@hms.harvard.edu (S.V.)
https://doi.org/10.1016/j.stem.2023.11.010

SUMMARY

Our understanding of pluripotency remains limited: iPSC generation has only been established for a few model species, pluripotent stem cell lines exhibit inconsistent developmental potential, and germline transmission has only been demonstrated for mice and rats. By swapping structural elements between Sox2 and Sox17, we built a chimeric super-SOX factor, Sox2-17, that enhanced iPSC generation in five tested species: mouse, human, cynomolgus monkey, cow, and pig. A swap of alanine to valine at the interface between Sox2 and Oct4 delivered a gain of function by stabilizing Sox2/Oct4 dimerization on DNA, enabling generation of high-quality OSKM iPSCs capable of supporting the development of healthy all-iPSC mice. Sox2/Oct4 dimerization emerged as the core driver of naive pluripotency with its levels diminished upon priming. Transient overexpression of the SK cocktail (Sox+Klf4) restored the dimerization and boosted the developmental potential of pluripotent stem cells across species, providing a universal method for naive reset in mammals.

INTRODUCTION

The discovery of induced pluripotent stem cells (iPSCs) by Takahashi and Yamanaka[1] has made enormous contributions to basic research, allowed new strategies for drug discovery, and provided a source for cell replacement therapy.[2] Pluripotent stem cells (PSCs) are unique in their ability to give rise to all tissues of the animal body; as such, they are the most developmentally potent cells we have in culture. The induction of pluripotency in somatic cells requires a complete epigenetic reset, which was once thought to be impossible.[3]

Oct4, Sox2, Klf4, and cMyc (OSKM)—all components of the Yamanaka cocktail—evolved to not only induce pluripotency in the inner cell mass (ICM) of the embryo[4,5] but also allow or even drive subsequent differentiation. Oct4, Sox2, and Klf4 (OSK) are pioneer transcription factors (TFs) capable of engaging silent chromatin; iPSC technology harnesses their pioneering

ability to rejuvenate somatic cells in vitro.[6,7] Oct4 stands out as the master regulator of the pluripotency network. Oct4 knockout in embryonic stem cells (ESCs) leads to a collapse of pluripotency.[8,9] Interestingly, downregulation, but not complete elimination, of Oct4 expression in ESCs leads to the opposite—stabilization of the pluripotency network, suggesting an additional role of Oct4 in differentiation.[10] Oct4 has been considered the only factor that cannot be replaced by other members of its family in iPSC generation.[11] However, it is endogenous Sox2 activation that signifies the completion of pluripotency induction.[12] Moreover, exogenous Oct4 causes a loss of developmental potential for OSKM versus SKM iPSCs,[13] suggesting that fine-tuning Oct4's functions might help to advance iPSC technology.

During mouse development, the future cell fate is biased already at the 4-cell stage, where high Sox2 expression and long-lived Sox2/Oct4 co-binding drive the emergence of the ICM.[14,15] Mice and humans have different degrees of Oct4

dependence when establishing pluripotency during early development: Oct4 knockout mouse blastocysts still develop a Nanog$^+$ ICM, whereas human OCT4-null blastocysts fail to do so.[16,17] Correspondingly, SKM induction is sufficient to induce pluripotency in mouse somatic cells,[13,18] but not in humans,[19] emphasizing the need to develop alternative strategies to improve the fidelity of non-murine reprogramming.

Oct4 cooperates with Sox2 to co-regulate most of its targets in pluripotent cells.[20] Sox2/Oct4 cooperativity is mediated by protein-protein interaction between their DNA-binding domains and by DNA allostery.[21] At the onset of reprogramming, when native sites are inaccessible, Oct4 and Sox2 often bind independently[7]; however, the sites engaged by both will most likely be opened.[22,23] Sox2/Oct4 heterodimerization, particularly on the canonical *HoxB1*-like *SoxOct* motifs, was shown to be essential for the induction and maintenance of pluripotency.[24] Sox17 can also cooperate with Oct4, but on compressed *SoxOct* motifs, controlling primitive endoderm and germline fates.[11,25–30] Jauch et al. discovered that a single residue swap between Sox17 and Sox2, glutamate to lysine at HMG box position 57 (Sox17^{E57K}), shifts its binding preference to the canonical *SoxOct* converting Sox17 into a pluripotency inducer.[30] Furthermore, the larger and more potent Sox17 C-terminus transactivator can enhance Sox2 function.[25,31,32] Here, we found that replacing Sox2 with Sox17^{E57K} in the reprogramming cocktail can rescue disabling Oct4 mutants and allows iPSC generation with somatic POU factors. We generated a library of chimeric Sox2-Sox17 TFs to find the structural elements of Sox17 responsible for this striking phenotype. The library screen allowed us to build an enhanced chimeric reprogramming factor that does not occur in nature. Our insights into the structure/function paradigm of Sox2 and Oct4 have major implications for understanding early development.

RESULTS

Defining the structural elements of Sox17 that enable iPSC generation

Oct4 (Pou5f1) is the only TF of the POU family that can induce pluripotency in mice and humans,[1,33,34] unlike other family members such as Oct1, Oct2, Oct6, and Brn4.[11,35,36] POU TFs exhibit different preferences for hetero- versus homo-dimerization.[22,35,37] In our search to find what makes Oct4 unique among POU factors, we studied its reprogramming ability in comparison with Brn4.[36] We discovered that Sox17^{E57K},[30] but not Sox2, can efficiently generate iPSCs in combination with Brn4 (Figure 1A). POU factors possess a DNA-binding POU domain, flanked by unstructured N- and C-terminal transactivator domains (NTD and CTD). The POU domain is bipartite, consisting of a POU-specific (POU$_S$) and POU-homeodomain (POU$_{HD}$) joined by a flexible non-conserved linker. The Oct4- but not Oct1-linker contains an alpha-helix at its N terminus.[38,39] Replacement of the Oct4-linker with those from other POU factors or the L80A mutation in the linker helix is detrimental for induction and maintenance of pluripotency.[38,40,41] Surprisingly, Sox17^{E57K} could also rescue the reprogramming ability of Oct4^{L80A} (Figures S1A and S1B).

Sox proteins consist of a DNA-binding HMG domain followed by an unstructured CTD (Figure 1B). We constructed a library of retroviral vectors carrying chimeric Sox2-Sox17 TFs, swapping the non-conserved regions of Sox2 with the respective regions of Sox17 (Figures 1B and 1C). The library was used to reprogram mouse embryonic fibroblasts (MEFs) carrying an Oct4-GFP reporter (Figures 1D and 1E); the volumes of viral supernatants were adjusted for equal expression (Figures S1C and S1D). Initial results pointed to residues 61–62 of Sox17 as the most crucial for rescuing Oct4^{L80A} (Figure 1E). The Sox17-CTD, 43–47, and 65–86 regions additionally boosted the reprogramming efficiency (Figures 1D and 1E). Further dissection determined that a single A61V swap in Sox2 allowed iPSC generation with Oct4^{L80A}, Brn4, Oct2, Oct6, and Brn2, but not with Oct1 (Figures 1F–1H). Replacing A61 with leucine (Sox2^{A61L}) performed better than wild-type (WT) Sox2 but worse than Sox2^{A61V} (Sox2^{AV}) (Figures 1F and 1G). Our best-performing chimeric factor, Sox2$^{43–47,61,65–86}$c17 (Sox2-17), which incorporates 17 residues of the Sox17-HMG and Sox17-CTD (Figures 1B and 1C), significantly improved the reprogramming efficiency, with 61V being the key residue (Figure 1G). cMyc, GATA factors, and the SV40 large T-antigen increase reprogramming efficiency by boosting cell proliferation,[13,47] but A61V has the opposite effect: four-factor induction with Sox2^{AV} or Sox2-17 resulted in significantly lowered cell proliferation compared with the respective A61 variants (Figures 1I and 1J). The repressive effect on cell proliferation might explain why A61V, although being able to rescue non-functional Oct4 mutants, does not by itself increase reprogramming efficiency with WT-Oct4. The efficiency boost comes from the synergy between A61V and the more potent Sox17-CTD (Figures 1D–1G).

Sox2^{AV} enhances Sox2/POU dimerization on canonical *SoxOct* motifs

Residue A61 of Sox2-HMG faces the Oct4-POU$_S$ when co-bound to a consensus *SoxOct* motif (Figure 2A). The extra methyl groups on valine make it more hydrophobic than alanine. Molecular dynamics simulations (MDSs) of the Sox2/Oct4 dimer on *Pou5f1* distal enhancer DNA element (*Oct4DE*) showed that A61V increases the average number of interactions between residue 61 and the POU$_S$ (Figure 2B). MDS of Oct4/Sox2 and Oct6/Sox2 versus the respective Sox2^{AV} dimers on *HoxB1* enhancer *SoxOct* DNA showed a similar increase in interactions (Figure S2A). In both sets of MDS, POU residue I21, conserved in Oct4 and Oct6, engaged V61 the most (Figures 2C and S2B).

Our MDS revealed a Sox2/Oct4 dimer configuration where HMG residues R50 and K57 form salt bridges with E82 and Q81 of the Oct4 linker (Figure 2A), similar to our previous report.[27] This arrangement involves both POU$_S$ and linker, hence the SL configuration, as opposed to the S configuration that involves only the POU$_S$ (Figure 2D). The SL configuration is Oct4 specific, as it was never observed for Oct6, which lacks negative charges in its linker (Figure 2D). The SL configuration dominated our Sox2/Oct4 simulations on *Oct4DE*, which were run with an AlphaFold-predicted Oct4 structure.[42] However, salt bridges between Sox2 and the Oct4-linker (residues E82 and E78) also occurred in simulations on *HoxB1* and *Nanog* regulatory DNAs, which were run with an Oct4 structure derived by crystallography[38] (Figures S2C and S2D; STAR Methods).

Modeling the Sox2/Oct4 dimer on the non-canonical *Fgf4* motif, where Sox and Oct sites are separated by a three
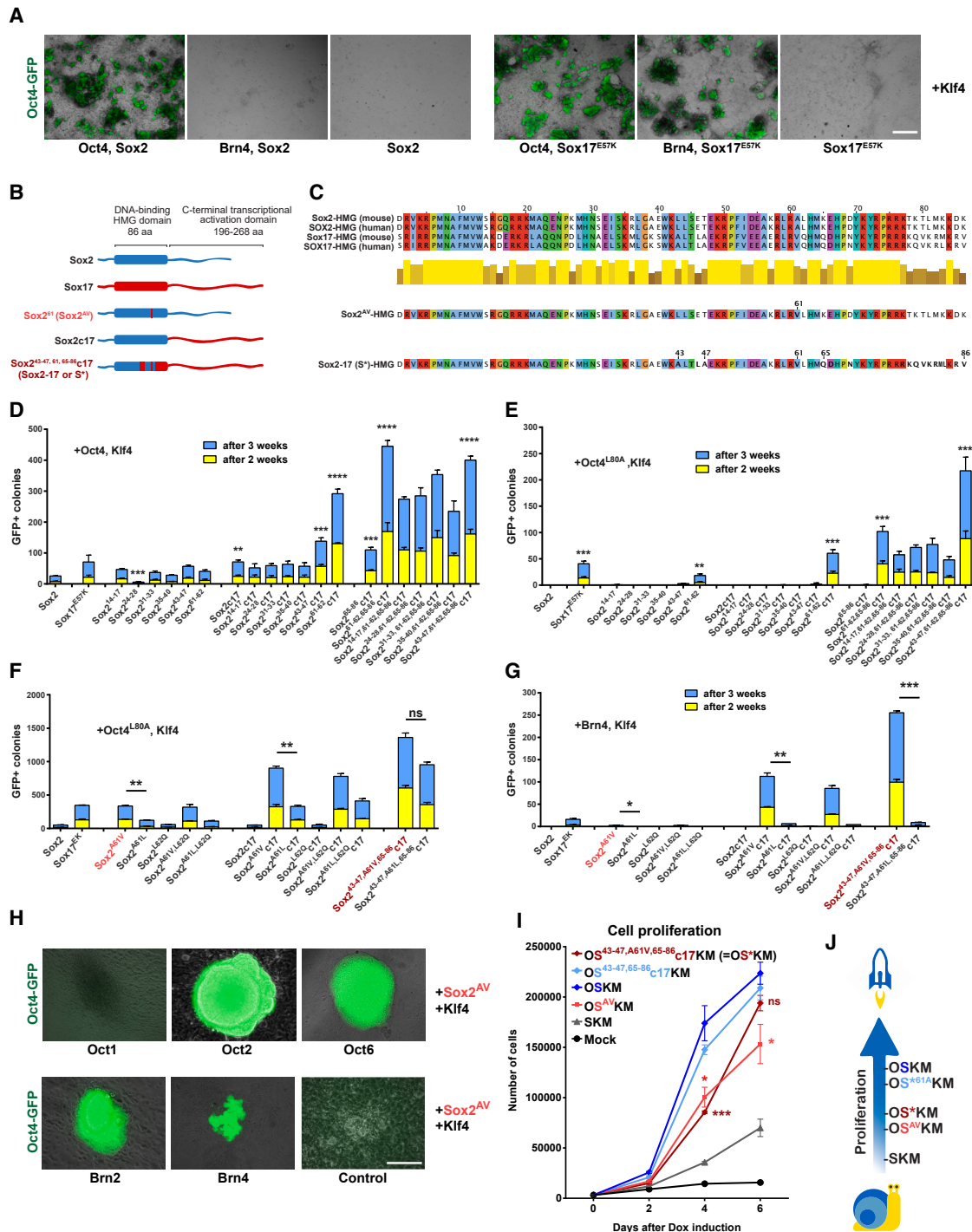
Figure 1. Reprogramming screen of Sox2-Sox17 chimeric TF library

(A) Bright-field and Oct4-GFP merged overview images showing retroviral reprogramming of Oct4-GFP (OG2) MEFs on 21 days post infection (scale bars, 1 mm).

(B) Schematic representation of Sox2 and Sox17 structures and chimeric TFs. Sequence from Sox2 in blue and Sox17 in red. Superscripts represent residues swapped from Sox17- to Sox2-HMG, e.g., Sox2$^{42-46}$ has residues 43–47 from Sox17, and Sox2c17 represents a complete replacement of Sox2-CTD with Sox17-CTD.

(C) Protein sequence alignment of DNA binding domains of mouse and human Sox2, Sox17, and the most crucial chimeric Soxes in this study.

(D–G) Reprogramming of OG2 MEFs by retroviral vectors carrying Klf4, Sox2/Sox17 chimeric TFs, and Oct4 (D), Oct4^{L80A} (E and F), or Brn4 (G).

(legend continued on next page)

base-pair gap,[43] revealed a Distant S (DS) configuration that involves Sox2's R75,[24] T78, and T80, but not A61 (Figure 2E).

We overexpressed FLAG-tagged Oct1, Oct2, Oct4, Oct6, Brn2, and Brn4 in HEK293 cells, adjusted for comparable expression (Figure S3A), and used the lysates for electromobility shift assays (EMSAs) on the *Nanog* promoter locus containing a canonical *SoxOct* motif.[44] Monomer binding was comparable among POU factors with the exception of Oct1, whereas Oct4 showed the strongest heterodimerization with Sox2. Both Sox2[AV] and Sox2-17 displayed stronger heterodimer bands with all tested POU TFs (Figure S3B).

Sox17 heterodimerizes on a compressed *SoxOct* DNA-controlling extraembryonic endodermal genes,[45] whereas compressed binding is disabled in the Sox17[E57K] mutant.[25,27,30] Consistently, neither Sox2[AV] nor Sox2-17 could dimerize with Oct4 on the compressed motif (Figures S3C and S3D).

We replaced all 17 residues of the Oct4-linker domain with poly-glycine linkers of different lengths (GL3-30) (Figure 3A). Such flexible linkers were detrimental for reprogramming with Sox2, but GL15-30 was rescued by Sox2[AV] (Figure 3A). We truncated Oct4 transactivators, both of which are crucial for reprogramming.[46] Neither Oct4ΔNTD nor Oct4ΔCTD could generate iPSCs when combined with Sox2 (Figures 3B, S3E, and S3F); Sox2[AV] and Sox2c17 could rescue Oct4ΔCTD, whereas Sox2[AV]c17, Sox17[EK], and Sox2-17 rescued both Oct4ΔCTD and Oct4ΔNTD (Figures 3B, S3E, and S3F). However, none of the chimeric Soxes rescued the deletion of the Oct4-POU$_S$, known to directly contact the Sox-HMG (Figure 3C). Sox2[AV] gave rise to a few iPSC colonies with Oct4ΔPOU$_{HD}$, verified by PCR genotyping and contribution to chimeric mice, including the germ line (Figures 3C and S3H–S3J). We conclude that Sox2[AV] could rescue the deletions of any Oct4 domain except for the POU$_S$, underlining the key role of Sox2/Oct4 dimerization in the induction of pluripotency.

We overexpressed Oct4 and Sox2 mutants in HEK293 cells (Figure 3D) and performed whole-cell lysate EMSAs. Monomer binding was similar between Sox2 and Sox2[AV]; however, A61V increased the dimerization with Oct4 on *Nanog*[44] and *HoxB1*[49] DNAs and partially rescued POU$_{HD}$ deletion (Figure 3E), in concordance with our reprogramming results (Figures 3C and S3G–S3J). We performed off-rate EMSAs by adding unlabeled DNA to the pre-formed Sox/Oct/DNA complex and loading samples over a time course. Sox monomer half-lives were comparable (Figure S3K), whereas both Sox2[AV] and Sox2-17 enhanced the heterodimer stability on *Oct4DE*[50] and *Nanog* elements, yet showed similar stability on the *Fgf4* motif[51] (Figure 3F), in line with our structural data (Figure 2E). A portion of Sox2/Oct4/*Oct4DE* dissociated immediately, although the remaining complex was long lived (Figure 3F), suggesting the presence of at least two Sox/Oct/DNA populations as in our MDS (Figures 2 and S2). We verified the whole-cell lysate results using purified proteins on *Nanog* and *Utf1*[52] DNAs (Figures 3G, S3L, and S3M). Sox2[AV] also increased the stability of heterodimers with

Oct4-linker mutants and Brn4 (Figure S3N). We conclude that the unique helical linker structure of Oct4 is functionally dispensable in the context of highly cooperative Soxes. This highlights the function of the Oct4-linker in dimerization with Sox2, likely through the SL configuration (Figures 2A, S2C, and S2D), explaining the Oct4-linker's key role in pluripotency.[38,40,41]

Oct4's pioneering function in development requires the ability to bind nucleosomal DNA,[7,53] posing a question about the role of POU subdomains in the process. Although a nucleosome does not hamper Sox2-alone binding,[54] it hinders canonical Oct4 binding because its POU$_S$ and POU$_{HD}$ engage opposite sides of DNA inevitably colliding with the histone core.[44,55–57] Thus, the reduced dependence on the POU$_{HD}$ in the presence of A61V could theoretically enhance heterodimer engagement of closed chromatin. We assembled reconstituted nucleosomes using the Widom 601 sequence with a *SoxOct* motif at the super-helical location (SHL) + 6, previously used to resolve the Sox2/Oct4/nucleosome.[56] Our EMSAs showed that A61V dramatically enhanced the stability of the Sox2/Oct4/nucleosome complex (Figure 4A).

We performed chromatin immunoprecipitation with sequencing (ChIP-seq) for MEF samples 48 h after doxycycline (Dox) induction of KS (tetO-Klf4-IRES-Sox2/Sox2[AV]) or OKS (tetO-Oct4/Oct6+tetO-Klf4-IRES-Sox2/Sox2[AV]). HOMER motif analysis[59] showed that all OKS samples were significantly enriched in *SoxOct* motifs (Figures S4A and S4B). Sox2 ChIP showed no significant difference for Sox2 and Sox2[AV] in KS samples and a relatively small difference in OKS samples (Figures 4B and 4C), suggesting that A61V does not change the Sox2 binding profile. However, Oct4 binding was significantly enhanced in the presence of Sox2[AV] (Figures 4B, 4C, and S4C). ChIP for both Oct4 and Oct6 showed an increased proportion of *SoxOct*-containing peaks in Sox2[AV] compared with Sox2 samples (Figure 4D), suggesting a genome-wide redistribution of POU binding. The enhanced Sox2[AV]/Oct4 dimer binding is demonstrated by the increased occupancy of Oct4 and Sox2[AV] at key naive pluripotency loci (*Klf2* and *Oct4DE*; Figure 4E). In line with our modeling and EMSA results (Figures 2E and 3F), the binding at the *Fgf4* locus remained unaffected (Figure 4E). Gene ontology analysis using GREAT[60] showed that differentially bound peaks in OKS[AV] samples were enriched in terms associated with early embryo development, the Wnt pathway, and negative regulation of cell proliferation, whereas OKS samples were enriched in terms related to activation of cell division through Hippo and MAPK pathways (Figure S4D).

In ESCs, Oct4 and Sox2 regulate most of their target genes cooperatively by binding *SoxOct* motifs.[61] At the beginning of the reprogramming process, the pluripotency genes are inaccessible, and the forcefully expressed Oct4 and Sox2 bind more independently, engaging thousands of non-native genomic loci.[7,22,23,48,62] Enhancing Sox2/Oct4 dimerization could potentially improve the reprogramming process, as cooperativity between TFs increases their specificity.[63] Indeed, already on day

---

(H) Representative phase-contrast and Oct4-GFP merged microscopy images of iPSCs colonies generated with retroviral vectors, control = S[AV]+K (no POU), scale bars, 200 μm.

(I) Cell proliferation assay.

(J) Summary of (I).

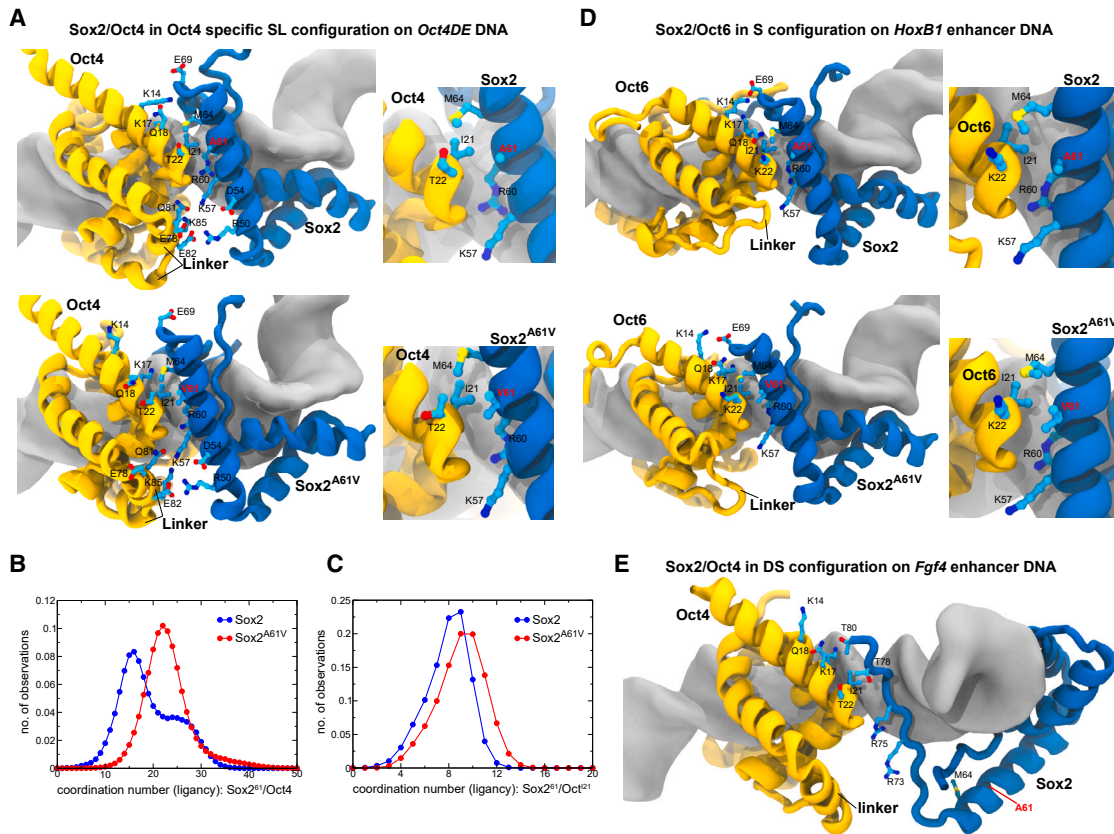Data in (D)–(G) and (I) represent mean ± SD; n = 3 biological replicates; Student's t test.

Figure 2. Molecular dynamic simulations reveal SL configuration of Sox2/Oct4

(A) Models of Sox2/Oct4 and Sox2^{AV}/Oct4 heterodimers on *Oct4DE* DNA in Oct4-specific SL configuration. The snapshots were captured from MDS in (B).

(B and C) MDS of Sox/Oct heterodimers on *Oct4DE* DNA. Plots show the number of contacts (ligancy) between HMG[48] and POU (B), or with Oct4[I21] (C). Detailed in STAR Methods.

(D) Models of Sox2/Oct6 and Sox2^{AV}/Oct6 heterodimers in S configuration on *HoxB1* enhancer DNA.

(E) Model of Sox2/Oct4 binding in DS configuration on *Fgf4* motif.

Only DNA-binding domains are shown in (A), (D), and (E).

2 of OKS induction, Sox2^{AV} engaged 511 of ESC-specific super-enhancers,[64] compared with 378 for Sox2 (Figure S4E). We performed TOBIAS footprinting analysis[65] using a publicly available assay for transposase-accessible chromatin using sequencing (ATAC-seq) datasets for ESCs versus MEFs.[58] Sox2/Oct4 footprints detected in ESC versus MEF represent the genomic loci to be opened during the reprogramming process. The occupancy of Sox2^{AV} in those key loci was slightly higher compared with Sox2, but Oct4 occupancy increased in OKS^{AV} samples (Figure 4F), suggesting a more robust ability of the Sox2^{AV}/Oct4 dimer to engage closed chromatin in early reprogramming (Figures 4A and S4E).

Stabilizing Sox2/Oct4 dimerization enhances the developmental potential of iPSCs

We performed tetraploid (4N) complementation assays to generate all-iPSC mice.[66–68] Mouse iPSC (miPSC) lines were generated using either lentiviral tet-inducible (pHAGE2-tetO) or episomal (pCXLE) vectors, both carrying polycistronic cassettes containing either Sox2 (OSKM) or Sox2^{AV} (OS^{AV}KM). Remarkably, all 10 tested OS^{AV}KM iPSC lines supported full-term devel-

opment of the aggregated embryos, whereas 3 out of 8 tested OSKM lines were incapable of supporting full-term development, echoing previous studies (Figure 4G; Table S1).[13,69,70] On average, OS^{AV}KM iPSCs gave rise to more than twice as many all-iPSC full-term pups as OSKM (Figure 4G; Table S1). OSKM all-iPSC mice rarely survive to adulthood[13,70,71]: of 25 pups born from 9 tetO-OSKM iPSC lines, none gave rise to adult all-iPSC mice (Table S1; Velychko et al.[13]). On the other hand, 4 out of 6 tetO-OS^{AV}KM lines gave rise to adult all-iPSC mice: of 68 tetO-OS^{AV}KM pups, 16 became healthy adults with 50% survival for the best-performing iPSC line (Figures 4G and 4H; Table S1). The tetO-OS^{AV}KM all-iPSC mice were fertile; the transgene inheritance was confirmed by PCR genotyping (Figure 4I). Episomal vectors deliver milder overexpression and give rise to overall better quality iPSCs, even in the presence of exogenous Oct4.[13] However, only 4.2% of transferred episomal OSKM all-iPSC embryos gave rise to adult mice compared with 22.2% for OS^{AV}KM iPSCs. The highest-quality episomal OS^{AV}KM iPSC line outperformed the highest-quality OSKM line: 43.3% versus 15.2% of transferred embryos gave rise to adult all-iPSC mice (Table S1).

(legend on next page)
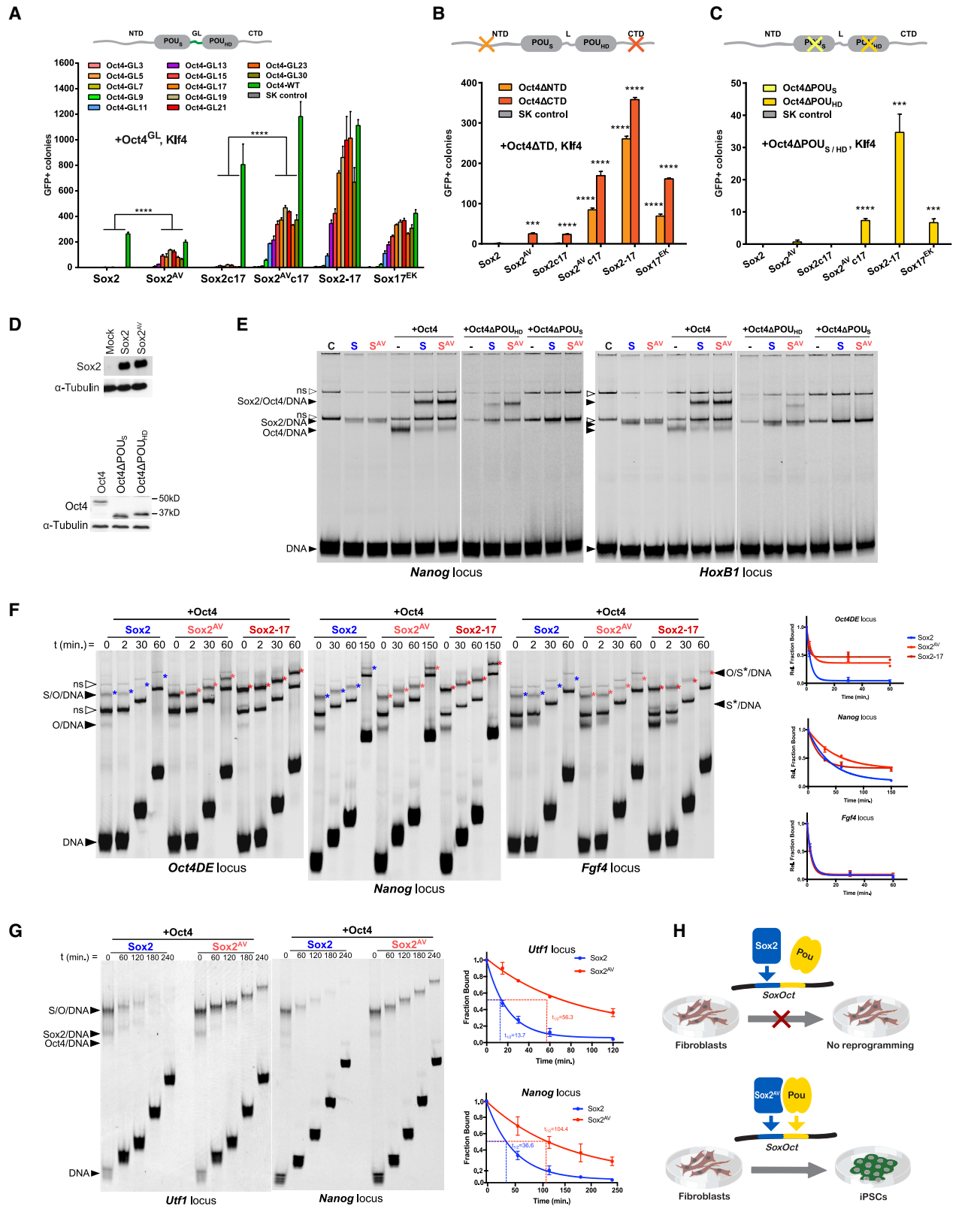
Cell Stem Cell
Article

Therefore, substituting a single residue of Sox2 enhances both Sox2/Oct4 dimerization capacity and the developmental potential of OSKM miPSCs.

Chimeric super-SOX enhances iPSC generation in five species

Sox2-17 (S* or super-Sox), which features A61V among other Sox17 elements, emerged as our most efficient chimeric reprogramming factor (Figures 1B and 1C), drawing interest for its practical applications. We cloned Sox2-17 into tet-inducible OSKM or SKM reprogramming cassettes and confirmed comparable levels of expression using RT-qPCR (Figure S5A). Time-course experiments with restricted Dox-induction (Figure 5A) showed that Sox2-17 enhanced the kinetics and efficiency of miPSC generation, shortening the minimal induction time from 3 days to just 24 h (Figure 5B). Clonally expanded 24 h-iPSCs lost methylation of Nanog and Pou5f1 promoters and acquired methylation of the fibroblast-specific Col1a1 promoter (Figure S5B), differentiated into all three germ layers in teratoma assays, contributed to chimeric mice, including the germ line, and successfully generated live-born all-iPSC pups in 4N complementation assays (Figures S5C–S5E; Table S1). When induced for just 3–4 days, OS*KM gave rise to 10–200 times more colonies than OSKM, depending on the quality of starting fibroblasts (Figures 5B and 5C). Sox2-17 could even generate two-factor miPSCs with Klf4, albeit with low efficiency (Figure S5F). S*K miPSC lines displayed mouse ESC-like (mESC) morphology, were verified by PCR genotyping, stained positive for Nanog and SSEA-1, and gave rise to three germ layers in a teratoma assay (Figures S5G–S5J). These data suggested that Sox2-17 requires shorter time, lower levels of expression, and a reduced number of additional factors to successfully induce pluripotency, which could be beneficial for the less efficient integration-free reprogramming methods. We generated episomal polycistronic OKS and OKS* vectors, carrying either Sox2 or Sox2-17, respectively, and confirmed the expression by western blot (Figure S5K). Sox2-17 enhanced episomal OKS MEF reprogramming by a striking 150 times, giving rise to high-quality miPSCs that could generate all-iPSC mice with up to 77% efficiency (Figures 5D and 5E; Table S1). Remarkably, all 10 tested OKS and OKS* iPSC lines gave rise to healthy adult mice with a survival rate similarly high for both Sox2 and Sox2-17 (Figure 5F; Table S1). This highlights that omitting Myc benefits the developmental potential of miPSC[70]; super-Sox offers a practical advantage by enhancing the OKS cocktail's efficiency.

Human iPSC (hiPSC) generation is far less efficient compared with mice; hence, deriving patient-specific hiPSC is often challenging. We tested a human version of SOX2-17 (Figure 1C) for reprogramming human fibroblasts using retroviral monocistronic OSKM (Figures 5G and 5H). SOX2-17 gave rise to 56 times more TRA1-60+ colonies compared with WT-SOX2: 8.9% versus 0.16% overall reprogramming efficiency. Self-replicating RNA-OKS*iG (VEE-OCT4-T2A-KLF4-E2A-SOX2-17-IRES-GLIS1) also generated 50 times more TRA1-60+ colonies than original OK-SiG[72] (Figure 5I). We used RNA-OKS*iG to generate iPSCs from dermal fibroblasts of patients with Parkinson's disease that could not be reprogrammed with the original OKSiG.[73]

SOX2-17, but not SOX17[E57K], could reprogram human cells even when combined with OCT4 alone, albeit with low efficiency; A61V was crucial for enabling the two-factor reprogramming (Figures 5J and 5K). OS*-hiPSCs showed hESC-like morphology, expressed NANOG and TRA1-81, and could differentiate into tissues of the three germ layers in teratoma assays (Figures 5L, 5M, and S5L).

We cloned SOX2[AV] and SOX2-17 into the episomal SK (pCXLE-SOX2-F2A-KLF4) vector,[74] replacing the F2A self-cleaving peptide with P2A to reduce the formation of a poly-protein.[36] Western blotting confirmed the correct cleavage of SOX and KLF4 (Figure S5M). The episomal SOX2-17-P2A-KLF4 combined with OCT4-shTP53 and L-MYC-F2A-LIN28 vectors (OS*KML) demonstrated improved reprogramming of aged human dermal fibroblasts compared with OSKML carrying WT-SOX2 (Figure 5N).

Although the generation of integration-free bona fide iPSCs is well established in mice and humans, the same cannot be said for many other species, including non-human primates (NHPs) and livestock. We tested SOX2-17 for reprogramming cynomolgus macaque fibroblasts.[75–77] OSKML failed to yield iPSCs despite multiple attempts, whereas OS*KML gave rise to alkaline phosphatases-positive (AP+) iPSC-like colonies that could be clonally expanded (Figure 5O). Although most hiPSC lines lose the episomes before passage 3, only 3 of 11 tested cynomolgus iPSC (ciPSC) lines lost the episomes; 2 of 3 integration-free lines had the correct chromosomal number, both displayed hiPSC-like morphology, expressed NANOG and OCT4, and differentiated into three germ layers in teratoma assays (Figures 5N, 5O, S5P, and S5Q).

We attempted to reprogram porcine and bovine fibroblasts using bFGF-based (StemFlex) media supplemented with XAV939, a Wnt inhibitor shown to support livestock ESC culture.[78] Episomal reprogramming using WT-SOX2 failed, whereas SOX2-17 efficiently generated AP+ colonies for both the pig and the cow that could give rise to clonal iPSC lines, which could be expanded beyond 12 passages (Figures 5P and 5Q). We established 12 bovine iPSC (biPSC) lines generated by OS*KML without p53 inhibition, which all lost the episomes by passage

---

Figure 3. Enhanced Sox/Oct cooperativity rescues non-functional POU factors in reprogramming

(A–C) OSK reprogramming of OG2 MEFs with monocistronic retroviral vectors carrying Oct4 domain deletion of linker (A), NTD or CTD (B), and POU_S or POU_{HD} (C).

(D) Western blot of whole-cell lysates from HEK293 used (E).

(E) EMSAs with HEK293 lysates on the Nanog promoter and HoxB1 enhancer DNA.

(F) Representative kinetic off-rate EMSAs with HEK293 lysates on Oct4DE, Nanog promoter, or Fgf4 enhancer DNA, asterisk = Sox/Oct/DNA.

(G) Kinetic off-rate EMSAs with purified proteins on Utf1 enhancer and Nanog promoter DNA. $t_{1/2}$ = ternary complex half-life. White arrowheads indicate nonspecific bands (ns) and black arrowheads indicate free DNA or DNA bound by Oct4, Sox2, or the heterodimer.

(H) Scheme showing the role of Sox/Oct dimerization in reprogramming.

Data in (A)–(C), (F), and (G) represent mean ± SD; n = 3 biological replicates (A–C) or experiments (F and G); Student's t test in (A)–(C).
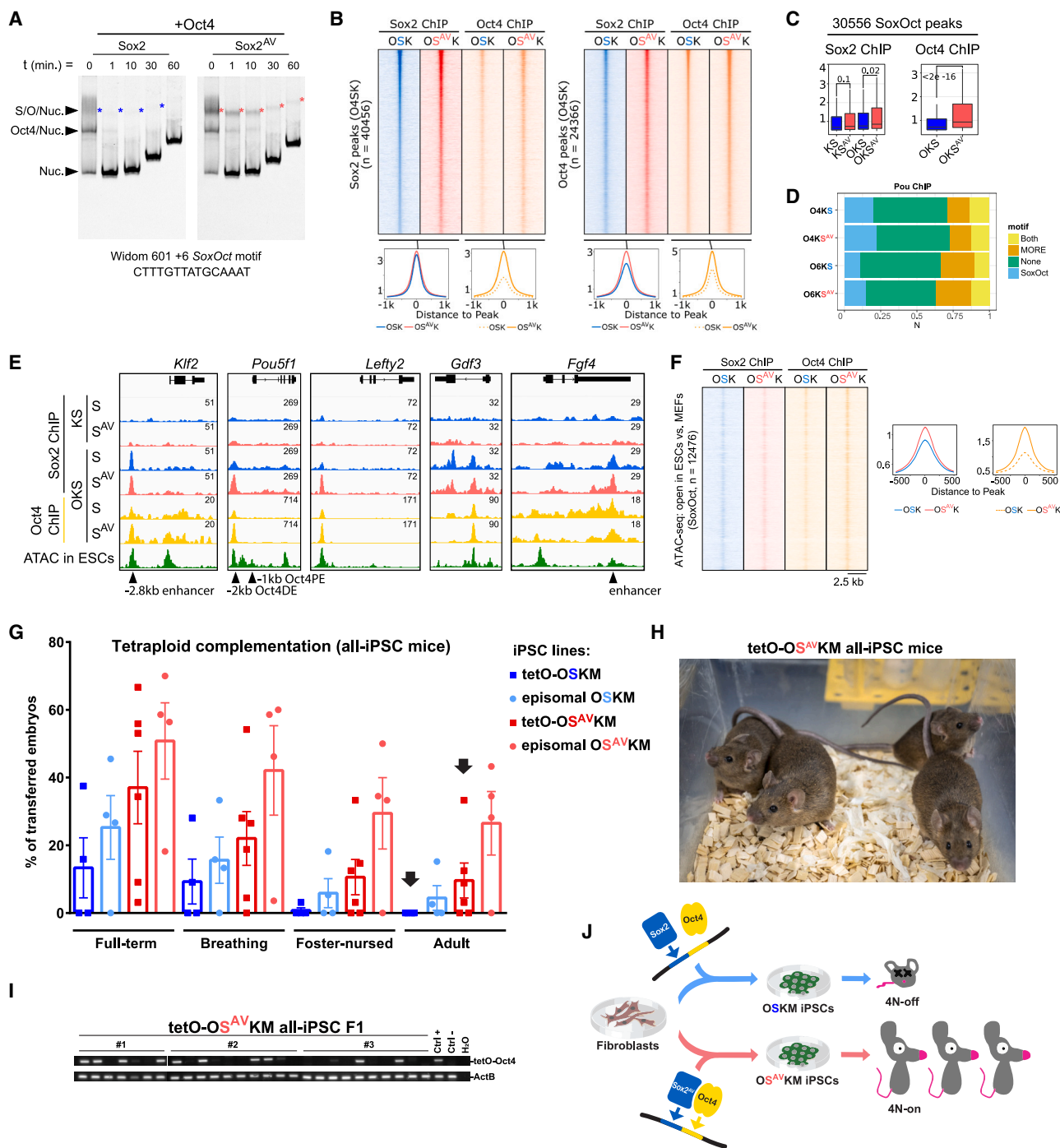
Figure 4. Highly cooperative Sox2^AV improves the developmental potential of mouse OSKM iPSCs

(A) Kinetic off-rate EMSAs with purified Sox2 or Sox2^AV co-bound with Oct4 on 601-SHL + 6 SoxOct nucleosome.

(B) Heatmaps and read pileup plots of ChIP-seq for MEF reprogramming samples at day 2 of Dox-induction.

(C) Boxplots of ChIP-seq peaks for OKS and KS reprogramming samples. The midline indicates the median, boxes indicate the upper and lower quartiles, and the whiskers indicate 1.5 times interquartile range. p values calculated using the unpaired Wilcoxon rank sum test.

(D) Fraction of binding sites containing SoxOct, MORE, both or none of the motifs in OKS reprogramming samples.

(E) Genome browser track of ChIP-seq peaks for selected loci.

(F) Heatmaps for ChIP-seq signals at the loci containing Sox2/Oct4 footprints in opened chromatin of mESCs versus MEFs, as determined by TOBIAS analysis of ATAC-seq data.[58]

(legend continued on next page)

6 (Figures S5R and S5S). biPSCs maintained ESC-like morphology, the correct number of chromosomes, and stained positive for SOX2 and OCT4 (Figures 5Q, S5T, and S5U). Thus, SOX2-17 allowed the generation of integration-free virus-free biPSCs with potential applications for cultivated beef and livestock gene editing.

We generated and characterized 30 clonal hiPSC lines derived from newborn foreskin (young, Y-iPSCs) and 56-year-old dermal (old, O-iPSCs) fibroblasts using episomal OSKML carrying WT-SOX2, SOX2^{AV}, or SOX2-17. All hiPSCs were integration-free with normal karyotypes (Figures S5V–S5X). Hierarchical clustering of both RNA-seq and reduced representation bisulfite sequencing (RRBS)[79] data showed that all hiPSC lines clustered far from fibroblasts and close to hESC lines (Figures 5R and 5S). The gene expression differences correlated more with the cell source rather than the SOX factors used. Loss of imprinting (LOI) is a common potentially cancerous irreversible epigenetic aberration afflicting iPSCs,[69,80–84] which correlates with poor developmental outcomes of all-iPSC embryos in 4N-complementation experiments.[13,69–71,80,85] We analyzed 23 differentially methylated regions (DMRs) represented in all samples and found that all lines including the original fibroblasts had different levels of LOI (Figure 5T). OS^{AV}KML-hiPSCs derived from young fibroblasts showed significantly lower levels of LOI compared with respective OSKML-iPSCs, whereas the differences between other hiPSCs were not significant. We conclude that highly cooperative Sox factors facilitate or enable iPSC generation in mammalian species (Figure 5U).

Sox2/Oct4 dimerization is at the core of naive pluripotency

The ESC derivative from mouse pre-implantation ICM mESCs and miPSCs represent the "naive" state; their proliferation in culture is dependent on LIF.[86] Naive mPSCs readily contribute to chimeric animals, and some lines are even capable of generating all-PSC mice. Conversely, PSCs from most other species, including humans, do not readily maintain the naive state and are typically stabilized in the "primed" state, which depends on FGF for proliferation. Mouse epiblast stem cells (mEpiSCs) derived from the post-implantation blastocyst are also primed—they share many characteristics with hPSCs, most importantly the low developmental potential.[87,88]

Oct4DE is active in naive but not primed PSCs in different species,[89–92] and both Sox2^{AV} and Sox2-17 increase the stability of the Sox2/Oct4 dimer on *Oct4DE* DNA (Figures 3F and 4E). We hypothesized that Sox2/Oct4 dimerization could be at the core of naive pluripotency.

We analyzed a published ATAC-seq dataset of time-course naive-to-primed transition samples generated by exposing mESCs to FGF.[93] The most significant changes occur between day 1 and day 2 of priming (Figure 6A). TOBIAS[65] footprinting analysis showed that the most depleted footprints between day 0 and day 1 were of Esrr and Klf factors (Figure 6B), con-

sistent with previous studies showing that Klf4 or Esrrb can reset mEpiSCs to the naive state.[94,95] More importantly, the day 1 → day 2 changes in chromatin landscape were dominated by the reduction of Sox/Oct and Sox footprints (Figure 6C).

We performed whole-cell lysate EMSAs using the *Nanog* element to measure the dimerization levels between Sox2 and Oct4 proteins endogenously expressed in different PSC lines: naive mESCs grown in KSR-LIF media, primed mEpiSCs carrying Oct4DE-GFP reporter (Gof18)[96–99] grown in FGF-based hESC media (StemFlex), mEpiSCs after naive reset grown in LIF or 2iLIF media,[100] and hiPSCs grown in hESC media (Figure 6D). The primed mEpiSCs and hiPSCs had significantly lower levels of Sox2/Oct4 dimer compared with mESCs, but the heterodimerization was restored in mEpiSCs after the naive reset by a brief exposure to MEK inhibitor, PD0325901, and sorting for Gof18$^+$. The heterodimerization was enhanced even further if the same cells were cultured in the presence of 2i (Figure 6D), which potentially points to the mechanism of the mouse naive media.[100] Primed cells of both species had more than twice lower ratio of Sox2/Oct4 dimer to Oct4 monomer binding compared with naive samples (Figure 6D). Antibody supershift confirmed the composition of EMSA bands (Figure 6E). The limited Sox2/Oct4 dimerization was due to lower Sox2 protein levels in primed cells, whereas there was no significant difference in Oct4 levels (Figures 6F and 6G). These data corroborate previous reports showing that mouse-primed cells have lower Sox2 expression compared with naive cells[99,101]; primed but not naive cells could even tolerate Sox2 knockout.[102]

mEpiSCs could be converted to the naive state by overexpression of Klf4[95]; however, lentiviral Klf4 alone could not reset human-primed iPSCs in KSR-LIF media (Figure 6H). Screening of different subsets of OSKM showed that SK (Sox2+Klf4) is the minimal cocktail that enables the generation of KLF17^{+}[103–106] hiPSCs. SK reset worked even in the absence of small molecule inhibitors (Figure 6H), but supplementing media with PD0325901 enhanced the efficiency of the reset (Figure S6A). Analogous to SKM miPSC generation,[13,18] combining Sox2 and Klf4 in a bicistronic vector proved crucial for the efficient naive reset of hiPSCs (Figure S6A).

We generated human episomal reprogramming plasmids mCherry-SK (pCXLE-mCherry-T2A-SOX2-P2A-KLF4) and mCherry-S*K (pCXLE-mCherry-T2A-SOX2-17-P2A-KLF4) to achieve a traceable integration-free naive reset. The episomal vectors were lipofected into mEpiSCs, and the mCherry$^+$/Gof18$^-$ cells were sorted on day 2 and plated on feeders in KSR-LIF media (Figures 6I and S6B). The majority of surviving cells formed dome-shaped colonies that were Gof18$^+$/mCherry$^-$ as early as day 4 after plating. We picked and clonally expanded 6 colonies for each cocktail. Both SK- and S*K-converted lines exhibited significantly higher Sox2/Oct4 dimerization than untransfected mEpiSCs, which correlated with increased Sox2 protein levels (Figures 6J, S6C, and S6D). SK-reset mEpiSC lines exhibited on average a 4-fold increase in heterodimer band

(G) Percentage of 4N-aggregated all-iPSC embryos. Data points represent means for each clonal iPSC line. Scale bars represent the mean ± SEM between all lines generated with the same cocktail and delivery method.

(H) Adult tetO-OS^{AV}KM all-iPSC mice (9 months).

(I) PCR genotyping of the progeny of all-iPSC mice derived from 3 tetO-OS^{AV}KM iPSC lines.

(J) Summary of (G)–(I).

(legend on next page)
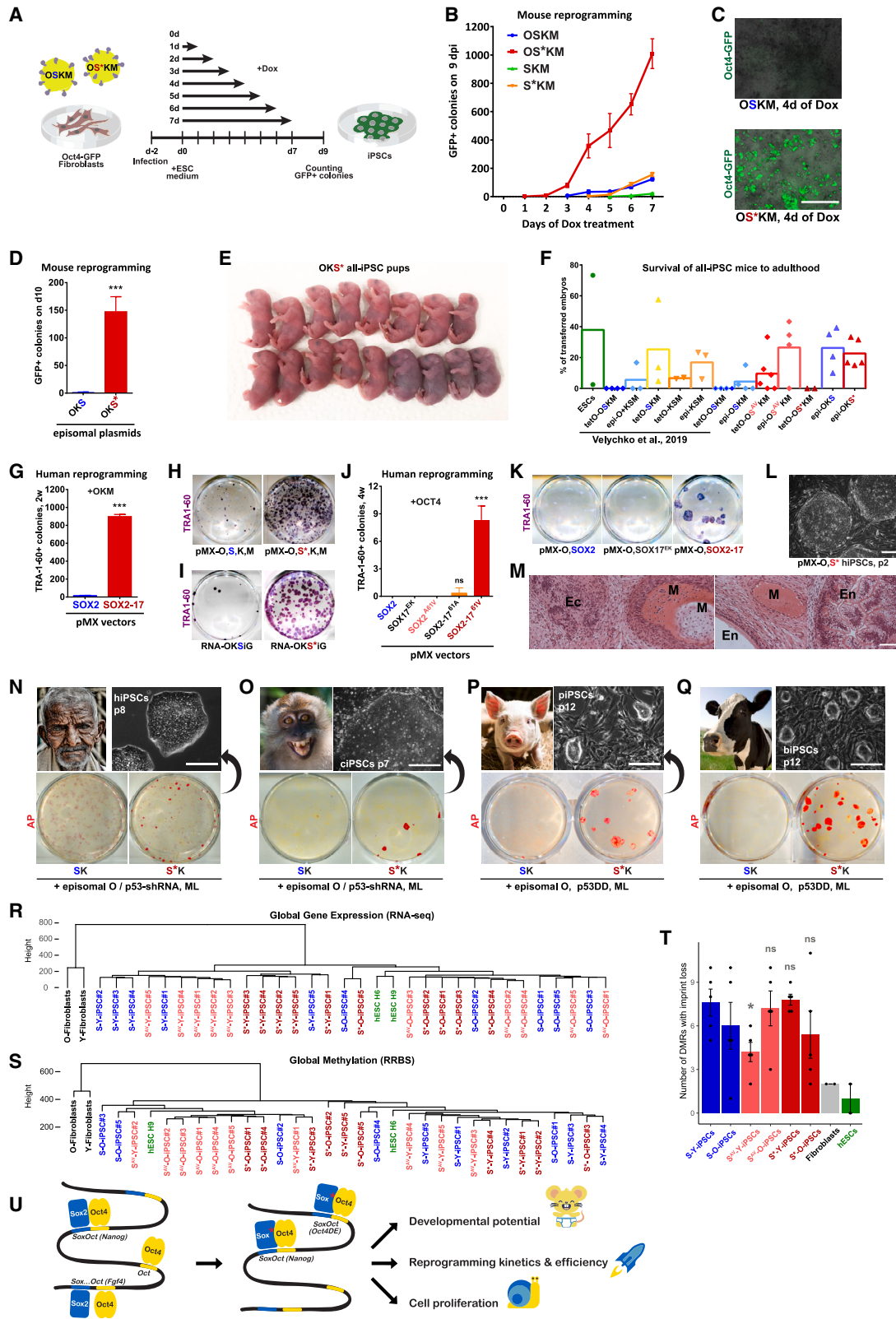
Cell Stem Cell
Article

intensity, compared with a 6-fold increase in S*K-reset mEpiSCs. The Sox2-17/Oct4 dimer band was not present in any of the S*K naive lines confirming that the episomal vectors were no longer expressed (Figure S6C). Compared with S*K-reset lines, the SK-reset naive lines had a significantly higher propensity to spontaneously lose Gof18+ status after passaging (Figures 6K and S6E), suggesting that S*K delivered a more stable naive reset.

Our data suggest that a decrease in Sox2/Oct4 dimerization is likely responsible for the downstream epigenetic changes that lead to diminished developmental potential upon priming of pluripotent cells in development and culture. Forced expression of Sox2 and Klf4 can efficiently reverse priming and convert mouse and human PSCs into the naive state (Figure 6L). Super-SOX, which exhibits enhanced cooperativity with POU factors, promotes both iPSC generation and naive reset, underscoring the key role of Sox2/Oct4 dimerization in naive pluripotency (Figures 6L and 6M). It would be interesting to investigate if the WT-Sox17+Klf4 cocktail can redistribute the Oct4 binding sites to compressed *SoxOct* motifs inducing primitive endoderm,[27,45] similarly to how Sox2+Klf4 induces the pre-implantation epiblast fate.

Episomal SK reset enhances the developmental potential of PSCs in three species

We co-nucleofected hiPSCs grown in primed media (StemFlex) with episomal mCherry-S*K and pCXWB-EBNA1 vectors and plated on feeders. After 48 h, the media was changed to human naive media (RSeT). By day 7, S*K-treated hiPSCs, but not control-nucleofected cells, generated dome-shaped colonies positive for human naive pluripotency markers SUSD2[107–109] and KLF17[103–106] (Figures 7A–7C). Day 7 SUSD2+ hiPSCs were mCherry− confirming the transgene-independent status of the generated naive cells. S*K gave rise to SUSD2+ and KLF17+ hiPSC colonies even in conventional feeder-free culture conditions in primed media (Figure 7D), a result not demonstrated for other naive cocktails.[92,110–114]

We performed RT-qPCR to assess the expression of key naive pluripotency genes (Figure 7E). S*K reset led to a significant up-regulation of *DNMT3L*, *KLF17*, and *ARGFX* in both primed and naive media. The naive media alone did not increase naive gene expression, except for a 6-fold upregulation of KLF4. Both fluorescence-activated cell sorting (FACS) and RT-qPCR data confirmed that the mCherry-S*K episome was eliminated from the cells by day 7, whereas the mCherry control plasmid persisted (Figures 7B and 7E), suggesting that S*K reset might trigger transgene silencing mechanisms as previously shown for mESCs[116] and the SKM cocktail.[13]

To test the developmental potential of our putative naive hiPSCs, we aggregated S*K-reset cells sorted for SUSD2+ at day 7 with mouse embryos at morula stage E2.5.[117] Astonishingly, S*K-reset hiPSCs marked with constitutive RFP expression were detected in the ICMs of the majority of cross-species aggregated embryos (Figure 7F). The cross-species chimerism was confirmed with co-staining of the chimeric embryos at E4.5 with human-specific SUSD2 and mouse-specific Oct4 antibodies. Human SUSD2+ cells were integrated into ICMs of 6 out of 11 embryos. In one case, the immunostaining indicated that S*K-reset hiPSCs took over the whole epiblast region (Figure 7G), which suggests that high levels of Sox2/Oct4 dimer might grant pluripotent cells an advantage in embryonic cell competition (Figure 7H).[115]

Initially, our OS*KML biPSCs failed to generate teratomas in severe combined immunodeficient (SCID) mice. Similarly, cultured bovine ESCs do not readily give rise to teratomas and lose to humans in cross-species cell competition (Figure 7H).[115] We injected control or S*K-reset biPSCs (Figure 7I) into opposite sides of the same mouse. A teratoma arose only from the S*K-reset sample, containing tissues representing all three embryonic germ layers (Figure 7J).

Finally, we nucleofected episomal mCherry or mCherry-S*K into a poor-quality naive female mESC line (C57BL/6J background) cultured in 2iLIF media. Emerging colonies were picked

Figure 5. Sox2-17 enhances iPSC generation in five species

(A) Schematic representation of time-course reprogramming experiment.

(B) Time-course reprogramming of Oct4-GFP MEFs; colonies were counted and imaged on day 9.

(C) Representative bright-field and Oct4-GFP merged images of day 4 samples from (B), scale bars, 2 mm.

(D) Episomal reprogramming of Oct4-GFP MEFs.

(E) Representative photo of all-iPSC pups derived from OKS* iPSC#1 from (D).

(F) Percentage of 4N-aggregated embryos that gave rise to healthy adult mice (survived at least 3 months), including our previous data for XY lines.[13] Scale bars represent mean.

(G) Reprogramming of human fetal fibroblasts with monocistronic retroviral (pMX) OSKM. TRA1-60+ colonies were counted at day 14.

(H and I) Representative whole-well scan of (G) or the same fibroblasts reprogrammed with self-replicating RNA (I).

(J) Two-factor reprogramming of human fibroblasts with monocistronic retroviral vectors. TRA1-60+ colonies were counted after 4 weeks.

(K) Representative whole-well scan of (J).

(L) Phase-contrast microscopy image of clonal hiPSC line generated in (J) and (K), scale bars, 200 μm.

(M) H&E staining of teratoma sections generated with OS* hiPSCs (ectoderm [Ec]: neural rosettes; mesoderm [M]: cartilage, bone, endothelium; endoderm [En]: gut and lung epithelium), scale bars, 70 μm.

(N–Q) Representative whole-well AP stainings for episomal reprogramming of (N) 56-year-old human male dermal fibroblasts on day 25, (O) Cynomolgus macaque fibroblasts on day 25, (P) porcine fetal fibroblasts on day 21, (Q) bovine fetal fibroblasts on day 21 after nucleofection. Representative phase-contrast images for clonal iPSC lines for each species are shown on top right, scale bars, 200 μm.

(R and S) Hierarchical clustering analysis of hESC and iPSCs derived from human newborn foreskin (young, Y) or 56-year-old human dermal (old, O) fibroblasts using episomal OSKML SOX2, SOX^AV, or SOX2-17 based on RNA-seq, TPM ≥ 1 (P), or RRBS (Q). Clustering was based on Euclidean distance.

(T) Comparison of LOI according to RRBS. Data points represent n of LOI for each clonal hiPSC line. Statistical significance calculated versus respective WT-SOX2 hiPSCs.

(U) A model of highly cooperative SOX in reprogramming.

Data in (B), (D), (G), (J), and (T) represent mean ± SD; n = 3 biological replicates for (B), (D), (G), and (J); Student's t test in (D), (G), (J), and (T).
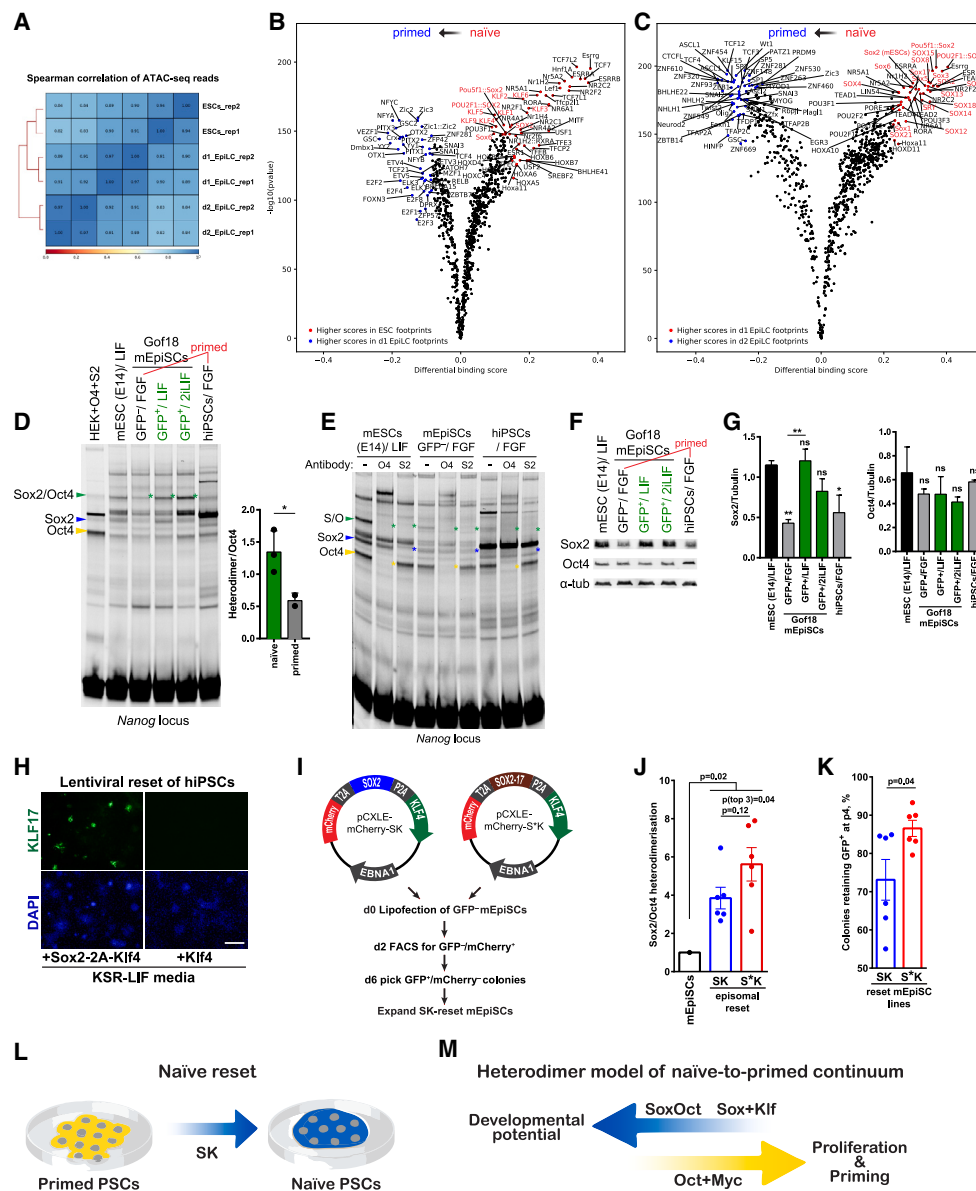
Figure 6. Sox2/Oct4 dimerization in naive versus primed pluripotency

(A) Spearman correlation of time-course ATAC-seq reads mESC samples undergoing priming.[93]

(B and C) TOBIAS footprinting analysis of (A) comparing mESCs versus day 1 EpiLC (B), and day 1 versus day 2 EpiLC samples (C).

(D) Representative EMSAs of whole-cell lysates of mESCs, Gof18⁻ mEpiSCs (E3), PD0325091-reset Gof18⁺ mEpiSCs cultured in KSR-LIF ± 2i, and hiPSCs on *Nanog* promoter DNA. All cells were grown on FSC-coated dishes without feeders.

(E) Supershift assay using anti-Oct4 and anti-Sox2 antibodies to confirm the identity of protein/DNA complexes in (D).

(F) Western blot of (D).

(G) Quantification of (F).

(H) Representative immunostaining for KLF17 of reset hiPSCs on day 6 after transduction with constitutive lentiviral vectors carrying Klf4 or Sox2-2A-Klf4, scale bars, 500 μm.

(I) Strategy for mEpiSC integration-free naive reset using episomal mCherry-SK.

(J) Quantification of Figure S6C EMSAs of clonal mEpiSC lines reset using episomal SK.

(K) Quantification of Gof18⁺ colonies from Figure S6E at passage 4, grown in KSR-LIF media on feeders. >100 colonies were quantified for each reset EpiSC line.

(L) Schematic protocol for naive reset.

(M) Heterodimer model of pluripotency states.

Data in (D) and (G) represent mean ± SD; n = 3 clonal lines (D) or biological replicates (G). Data points in (J) and (K) represent means for each clonal reset EpiSC line, bars represent mean ± SEM for all lines; n = 6 clonal lines. Student's t test in (D), (G), (J), and (K).
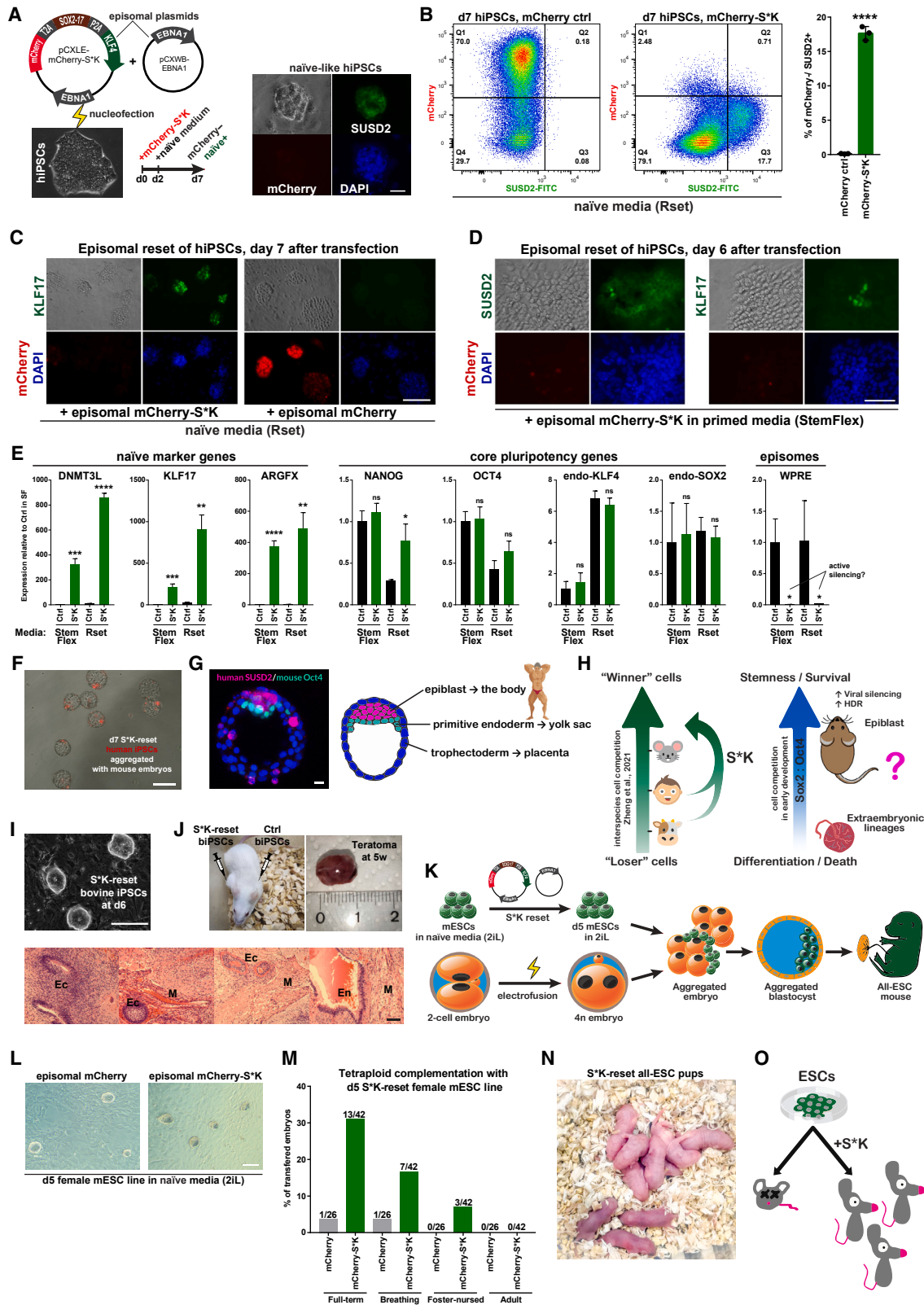
(legend on next page)

at day 5 and used for 4N complementation (Figures 7K and 7L). Remarkably, the S*K-reset cells generated 8 times more full-term all-ESC pups compared with the control (Figures 7M–7O; Table S1). Three S*K-reset all-ESC pups survived foster nursing, whereas the only control all-ESC pup died shortly after birth. The SKM cassette,[13] particularly when containing Sox2[AV], might further improve the naive reset, given the outstanding developmental potential of OS[AV]KM iPSCs (Figure 4 and Table S1).

The *in vivo* evidence for enhanced development potential in three species presented in this section, most importantly the birth of S*K-reset all-ESC animals, argues in favor of our proposed "heterodimer model" of pluripotency continuum (Figure 6L).

DISCUSSION

iPSC technology struggles with inefficiency and widely variable quality of the produced cell lines.[69,80,81,118] Some alternative reprogramming cocktails could improve the developmental potential of miPSCs, but they also decreased the reprogramming efficiency[13,70,71] and, consequently, failed to reprogram human cells that possess stronger epigenetic barriers.[19] To date, only the generation of all-iPSC mice has been reported,[66,67,119] and germline competence has only been demonstrated for mouse (both sexes) and male rat iPSCs,[120] highlighting the limitations of current technology. Here, we combined structural elements of Sox2 and Sox17 to build a chimeric super-Sox that enhanced reprogramming in five species: mouse, human, cynomolgus macaque, cow, and pig. The key point mutation, A61V, which stabilized Sox/Oct dimer on DNA, increased the developmental potential of OSKM miPSCs, as evidenced by higher rates of full-term development and survival of all-iPSC mice.

Oct4 functions independently of Sox2 to drive proliferation.[121,122] Notably, the cocktails[13,70] and culture interventions[123,124] that yield higher-quality iPSCs also reduce cell proliferation during reprogramming, suggesting that limiting proliferation is beneficial. This can be achieved by enhancing Sox/Oct dimerization, as in OS[AV]KM reprogramming; increasing Sox2:Oct4 ratio, as in SKM reprogramming[13] and Oct4 heterozygous knockout ESCs[10]; and omitting or reducing Myc, as in OKS[70] or OSKM compared with OKSM cassette,[69] and Myc-depleted ESCs.[125]

Mice are the only species, for which naive PSCs have been stabilized in culture without the use of small molecule inhibitors.[126] Mice likely evolved (or preserved) the unusual stability of their naive pluripotency fate to enable a blastocyst-stage embryonic arrest, known as diapause.[127] The developmental potential of mESCs and their increased capacity for homologous recombination repair (HDR) has allowed for unprecedented genetic engineering of this species.[128–130]

Supplementing culture media with certain small molecule inhibitors, alone or combined with TF-based cocktails, could generate naive-like PSCs for humans, NHPs, and other non-rodents, but the reported cells lack the functionality of mESCs—most importantly the ability to contribute to normal development and to give rise to functional germ line.[91,92,110,131–135] Long-term culture in naive media leads to epigenetic abnormalities and loss of germline competency for both humans and mice,[81,136,137] whereas a short exposure during reprogramming could be beneficial for hiPSC quality.[124] Contrary to mESCs, which exclusively contribute to the epiblast, chemically reset naive hESCs can also contribute to the trophectoderm,[138,139] which could be attributed to the low levels of Sox2/Oct4 dimerization.[106,140,141] The OSKM cocktail can induce naive pluripotency from somatic cells in mice[66,67] and humans.[124,142] In particular, the role of Klf4 in naive pluripotency has been described for both species.[111] Here, we showed that a subset of Yamanaka's cocktail containing Sox2 and Klf4 could induce naive reset in both mouse and human PSCs even in the absence of small molecule inhibitors. SK-reset links iPSC quality to the naive-primed continuum and explains the enhanced developmental potential of SKM miPSCs.[13] Episomal S*K reset improved the developmental potential in humans (evidenced by cross-species embryo aggregations), cows (generating teratoma-capable biPSCs), and mice (boosting all-ESC animal production). The *in vivo* evidence for naive reset in three species supports our proposed heterodimer model of a naive-to-primed pluripotency continuum, which elucidates the roles of Yamanaka factors: high levels of Sox2 and Klf4 expression and Sox2/Oct4 dimerization promote the naive state, whereas decreased Sox2 reduces the heterodimerization, and when coupled with excess Oct4 and Myc, promotes cell proliferation and priming.

Figure 7. Episomal S*K reset of human, bovine, and mouse PSCs

(A) Method for integration-free naive reset using pCXLE-mCherry-S*K. Immunostaining of day 7 S*K-reset hiPSCs stained for SUSD2, scale bars, 50 μm.

(B) FACS for SUSD2 of day 7 reset hiPSCs. Error bars represent SD; n = 3; Student's t test.

(C and D) Immunostaining of day 7 reset hiPSCs stained for KLF17 (RSeT media starting from day 2, C), and day 6 S*K-reset hiPSCs stained for SUSD2 and KLF17 (StemFlex media at all times, D), scale bars, 50 μm.

(E) RT-qPCR analysis of bulk day 7 reset samples. Expression normalized to *GAPDH*.

(F) Cross-species human/mouse morula aggregation with day 7 SUSD2+ hiPSCs marked by constitutive RFP, reset with pCXLE-S*K (no mCherry), E4, scale bars, 100 μm.

(G) Immunostaining of (F) with human-specific SUSD2 and mouse-specific Oct4 antibodies, scale bars, 10 μm.

(H) Model of interspecies and early development cell competition based on Zheng et al.[115] and our data.

(I) Representative phase-contrast image of day 6 S*K-reset biPSCs, scale bars, 200 μm.

(J) Teratoma generated by S*K-reset biPSCs (ectoderm [Ec]: neural rosettes, epidermis, squamous epithelial cells; mesoderm [M]: smooth muscles, connective tissue; endoderm [En]: gut epithelium), scale bars, 100 μm.

(K) Schematic representation of 4N-complementation experiment with S*K-reset mESCs.

(L) Representative phase-contrast images of day 5 S*K-reset mESCs, scale bars, 200 μm.

(M) Percentage of 4N-aggregated embryos generated in (K) and (L). Scale bars represent mean survival of transferred embryos; numbers are shown on top.

(N) All-iPSC pups generated by S*K-reset female mESCs.

(O) Summary of (K)–(N). Data in (B) and (E) represent mean ± SD; n = 3 biological replicates; Student's t test.

Cell Stem Cell
Article

CellPress
OPEN ACCESS

The evolutionary tree of animals suggests that of the Sox2/Oct4 couple, in the beginning, there was Sox2. Key Sox2 residues, such as R50 and K57, are already present in sponges, where SoxB TFs control early embryogenesis.[143] A61 is conserved in hydrozoans, where SoxB genes are expressed in stem cells that give rise to neuroectoderm (default) and germ line (when stressed).[144–146] POU5 factors emerged much later in the evolutionary tree—it is an innovation of vertebrates,[147,148] where POU5 TFs cooperate with Sox2 to control early development.[14,149] The Oct4-linker is not directly involved in DNA binding; however, it is important for reprogramming to pluripotency and for normal development.[38,40,41] Here, we found that Oct4-linker mutations reduce stability of the Sox2/Oct4 dimer on DNA, but $Sox2^{AV}$ could rescue linker mutants' ability to heterodimerize and to induce pluripotency. Likewise, $Sox2^{AV}$ enabled heterodimerization and reprogramming with tissue-specific POU factors. Our MDS revealed that the negatively charged Oct4-linker residues form salt bridges with positively charged R50 and K57 of Sox2. Although the linker is the least conserved POU subdomain, its negative charges are already present in the POU5 factor of jawless hagfish.[147] It has been suggested that two distinct POU5 factors that still exist in many vertebrates could support either naive or primed pluripotency,[147] possibly through their differential ability to cooperate with Sox2. Our work demonstrates that the most significant feature that distinguishes Oct4 from other POU factors is its ability to form a stable heterodimer with Sox2 that had already been in control of early embryogenesis in lower animals.

The heterodimer model of reset presented here aligns with other studies placing Sox2 at the top of the pluripotency hierarchy.[12,14,22,23,140,150–152] In early animal development, unidirectionality is likely achieved by a negative feedback loop limiting the return to a high-Sox state.[153] Interestingly, female mPSCs have lower developmental potential compared with male.[154,155] Our model suggests that the reason for higher developmental pluripotency in male lines could be the expression of sex-determining region of Y (Sry),[156] which has a Sox2-like DNA-binding motif. The "high-Sox" hypothesis could also explain the increased survival of male versus female embryos in humans and other mammals[157,158] and the higher occurrence of Sox-driven cancers in men versus women.[159] The abundance of Sox footprints in the open chromatin of naive versus primed cells suggests a more general developmental trend, where various ratios of Sox factors and their partners predispose stem cells toward specific lineages.

Our engineered super-SOX factor harnesses the reprogramming powers from naturally evolved structural elements of two major development regulators, Sox2 and Sox17. Even more efficient reprogramming factors could potentially be built by means of rational engineering and directed evolution.[160,161] Our data suggest that enhancing cooperativity between key co-factors should be one of the goals of future designers.

Limitations of the study

A small number of cells from the ICM can contribute disproportionally to animal development.[162] Thus, the developmental potential of a given PSC line may be determined by a Sox2/Oct4-high subpopulation rather than the average measured by ATAC-seq, EMSA, and western blot experiments. Further studies are needed to characterize the SK-reset naive PSCs and address the posttranslational modifications and other mechanisms regulating Sox2/Oct4 dimerization. Our current episomal reset protocol produces naive PSCs transiently—between days 4 and 7, requiring their use in downstream applications before re-priming occurs. A culture media supporting long-term maintenance of transgene-independent non-murine naive PSCs with high heterodimer levels remains to be formulated.

For this study, we generated OSKML hiPSC lines using the construct carrying shRNA against TP53,[74] which knocks down the main tumor suppressor boosting cell proliferation. TP53 knockdown is likely detrimental to the iPSC quality and could have caused LOI in our hiPSCs.

We cannot exclude that a highly cooperative Sox or excess of Sox2 may participate in the developmental reset in ways beyond enhancing Sox/Oct dimerization, e.g., by remodeling the epigenome through recruiting the aging antagonist Parp1[163] or silencing retroviral elements.[13]

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice
  - Primary cells
  - Cell lines
  - Microbe strains
- METHOD DETAILS
  - Vector construction
  - iPSC generation and characterization
  - Tetraploid (4N) complementation assay
  - Lentiviral naïve reset of human iPSCs
  - Integration-free naïve reset of mouse epiblast stem cells
  - Integration-free naïve reset of human PSCs using episomal vectors
  - Mammalian cell overexpression and whole-cell lysate (WCL) generation
  - Western blot analysis
  - Insect cell expression and protein purification
  - Electrophoretic mobility shift assays (EMSAs)
  - Nucleosome assembly
  - Molecular Dynamics Simulation (MDS)
  - NGS and bioinformatic analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.stem.2023.11.010.

AUTHOR CONTRIBUTIONS

S.V. conceived the study, designed and performed most of the experiments, interpreted the results, and wrote the manuscript with input from other authors. H.R.S. advised on the project, raised funding, and provided a collaborative environment. C.M.M. established and performed the biochemistry, interpreted the results, and edited the manuscript. V.C. advised on the project and performed and interpreted MDS. G.W. performed and interpreted 4N aggregations. V.M. performed ChIP-seq. V.M. and R.J. analyzed ChIP-seq. T.V. helped with cloning and reprogramming. G.K. analyzed RRBS. Y.M.-L. performed episomal hiPSC reset, imaged biPSC-teratoma, and interpreted the results. Y.M.-L., R.F., and I.B. performed human/mouse chimeric embryo aggregations, and G.M.C. advised on the project.

DECLARATION OF INTERESTS

REFERENCES

1. Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

2. Yamanaka, S. (2020). Pluripotent stem cell-based cell therapy—promise and challenges. Cell Stem Cell *27*, 523–531.

3. Waddington, C.H. (1957). The Strategy of the Genes: a Discussion of Some Aspects of Theoretical Biology (Allen & Unwin).

4. Wu, G., and Schöler, H.R. (2014). Role of Oct4 in the early embryo development. Cell Regen. *3*, 7.

5. Nakatake, Y., Fukui, N., Iwamatsu, Y., Masui, S., Takahashi, K., Yagi, R., Yagi, K., Miyazaki, J., Matoba, R., Ko, M.S.H., et al. (2006). Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells. Mol. Cell. Biol. *26*, 7772–7782.

6. King, H.W., and Klose, R.J. (2017). The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. eLife *6*, 1–24.

7. Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994–1004.

8. Nishimoto, M., Miyagi, S., Yamagishi, T., Sakaguchi, T., Niwa, H., Muramatsu, M., and Okuda, A. (2005). Oct-3/4 maintains the proliferative embryonic stem cell state via specific binding to a variant octamer sequence in the regulatory region of the UTF1 locus. Mol. Cell. Biol. *25*, 5084–5094.

9. Niwa, H., Masui, S., Chambers, I., Smith, A.G., and Miyazaki, J. (2002). Phenotypic complementation establishes requirements for specific POU domain and generic transactivation function of Oct-3/4 in embryonic stem cells. Mol. Cell. Biol. *22*, 1526–1536.

10. Karwacki-Neisius, V., Göke, J., Osorno, R., Halbritter, F., Ng, J.H., Weiße, A.Y., Wong, F.C.K., Gagliardi, A., Mullin, N.P., Festuccia, N., et al. (2013). Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog. Cell Stem Cell *12*, 531–545.

11. Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat. Biotechnol. *26*, 101–106.

12. Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell *150*, 1209–1222.

13. Velychko, S., Adachi, K., Kim, K.-P.P., Hou, Y., MacCarthy, C.M., Wu, G., and Schöler, H.R. (2019). Excluding Oct4 from Yamanaka cocktail unleashes the developmental potential of iPSCs. Cell Stem Cell *25*, 737–753.e4.

14. White, M.D., Angiolini, J.F., Alvarez, Y.D., Kaur, G., Zhao, Z.W., Mocskos, E., Bruno, L., Bissiere, S., Levi, V., and Plachta, N. (2016). Long-lived binding of Sox2 to DNA predicts cell fate in the four-cell mouse embryo. Cell *165*, 75–87.

15. Goolam, M., Scialdone, A., Graham, S.J.L., MacAulay, I.C., Jedrusik, A., Hupalowska, A., Voet, T., Marioni, J.C., and Zernicka-Goetz, M. (2016). Heterogeneity in Oct4 and Sox2 targets biases cell fate in 4-cell mouse embryos. Cell *165*, 61–74.

16. 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. Nat. Cell Biol. *15*, 1089–1097.

17. Fogarty, N.M.E., McCarthy, A., Snijders, K.E., Powell, B.E., Kubikova, N., Blakeley, P., Lea, R., Elder, K., Wamaitha, S.E., Kim, D., et al. (2017). Genome editing reveals a role for OCT4 in human embryogenesis. Nature *550*, 67–73.

18. An, Z., Liu, P., Zheng, J., Si, C., Li, T., Chen, Y., Ma, T., Zhang, M.Q., Zhou, Q., and Ding, S. (2019). Sox2 and Klf4 as the functional core in pluripotency induction without exogenous Oct4. Cell Rep. *29*, 1986–2000.e8.

19. Kim, K.P., Choi, J., Yoon, J., Bruder, J.M., Shin, B., Kim, J., Arauzo-Bravo, M.J., Han, D., Wu, G., Han, D.W., et al. (2021). Permissive epigenomes endow reprogramming competence to transcriptional regulators. Nat. Chem. Biol. *17*, 47–56.

20. Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., 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.

21. Merino, F., Bouvier, B., and Cojocaru, V. (2015). Cooperative DNA recognition modulated by an interplay between protein-protein interactions and DNA-mediated allostery. PLoS Comput. Biol. *11*, e1004287.

22. Malik, V., Glaser, L.V., Zimmer, D., Velychko, S., Weng, M., Holzner, M., Arend, M., Chen, Y., Srivastava, Y., Veerapandian, V., et al. (2019). Pluripotency reprogramming by competent and incompetent POU factors uncovers temporal dependency for Oct4 and Sox2. Nat. Commun. *10*, 3477.

23. Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative binding of transcription factors orchestrates reprogramming. Cell *168*, 442–459.e20.

24. Tapia, N., MacCarthy, C., Esch, D., Gabriele Marthaler, A., Tiemann, U., Araúzo-Bravo, M.J., Jauch, R., Cojocaru, V., and Schöler, H.R. (2015). Dissecting the role of distinct OCT4-SOX2 heterodimer configurations in pluripotency. Sci. Rep. *5*, 13533.

25. Aksoy, I., Jauch, R., Eras, V., Chng, W.B.A., Chen, J., Divakar, U., Ng, C.K.L., Kolatkar, P.R., and Stanton, L.W. (2013). Sox transcription factors require selective interactions with oct4 and specific transactivation functions to mediate reprogramming. Stem Cells 31, 2632–2646.

26. Ng, C.K.L., Li, N.X., Chee, S., Prabhakar, S., Kolatkar, P.R., and Jauch, R. (2012). Deciphering the Sox-Oct partner code by quantitative cooperativity measurements. Nucleic Acids Res. 40, 4933–4941.

27. Merino, F., Ng, C.K.L., Veerapandian, V., Schöler, H.R., Jauch, R., and Cojocaru, V. (2014). Structural basis for the SOX-dependent genomic redistribution of OCT4 in stem cell differentiation. Structure 22, 1274–1286.

28. Irie, N., Weinberger, L., Tang, W.W.C., Kobayashi, T., Viukov, S., Manor, Y.S., Dietmann, S., Hanna, J.H., and Surani, M.A. (2015). SOX17 is a critical specifier of human primordial germ cell fate. Cell 160, 253–268.

29. Wang, X., Veerapandian, V., Yang, X., Song, K., Xu, X., Cui, M., Yuan, W., Huang, Y., Xia, X., Yao, Z., et al. (2021). The chromatin accessibility landscape reveals distinct transcriptional regulation in the induction of human primordial germ cell-like cells from pluripotent stem cells. Stem Cell Rep. 16, 1245–1261.

30. Jauch, R., Aksoy, I., Hutchins, A.P., Ng, C.K.L., Tian, X.F., Chen, J., Palasingam, P., Robson, P., Stanton, L.W., and Kolatkar, P.R. (2011). Conversion of Sox17 into a pluripotency reprogramming factor by reengineering its association with Oct4 on DNA. Stem Cells 29, 940–951.

31. Hu, H., Ho, D.H.H., Tan, D.S., MacCarthy, C.M., Yu, C.H., Weng, M., Schöler, H.R., and Jauch, R. (2023). Evaluation of the determinants for improved pluripotency induction and maintenance by engineered SOX17. Nucleic Acids Res. 51, 8934–8956.

32. Weng, M., Hu, H., Graus, M.S., Tan, D.S., Gao, Y., Ren, S., Ho, D.H.H., Langer, J., Holzner, M., Huang, Y., et al. (2023). An engineered Sox17 induces somatic to neural stem cell fate transitions independently from pluripotency reprogramming. Sci. Adv. 9, eadh2501.

33. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 861–872.

34. Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science 318, 1917–1920.

35. Jerabek, S., Ng, C.K., Wu, G., Arauzo-Bravo, M.J., Kim, K.P., Esch, D., Malik, V., Chen, Y., Velychko, S., MacCarthy, C.M., et al. (2017). Changing dimerization preferences converts Oct6 into a pluripotency inducer. EMBO Rep. 18, 319–333.

36. Velychko, S., Kang, K., Kim, S.M., Kwak, T.H., Kim, K.P., Park, C., Hong, K., Chung, C.H., Hyun, J.K., MacCarthy, C.M., et al. (2019). Fusion of reprogramming factors alters the trajectory of somatic lineage conversion. Cell Rep. 27, 30–39.e4.

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

38. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589.

39. Reményi, A., Lins, K., Nissen, L.J.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 Dev. 17, 2048–2059.

40. Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., Zaret, K.S., Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., et al. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell 161, 1–14.

41. Aksoy, I., Jauch, R., Chen, J., Dyla, M., Divakar, U., Bogu, G.K., Teo, R., Leng Ng, C.K., Herath, W., Lili, S., et al. (2013). Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. EMBO J. 32, 938–953.

42. Kim, K.P., Wu, Y., Yoon, J., Adachi, K., Wu, G., Velychko, S., MacCarthy, C.M., Shin, B., Röpke, A., Arauzo-Bravo, M.J., et al. (2020). Reprogramming competence of OCT factors is determined by transactivation domains. Sci. Adv. 6, eaaz7364.

43. Esch, D., Vahokoski, J., Groves, M.R., Pogenberg, V., Cojocaru, V., vom Bruch, H., Han, D., Drexler, H.C., Araúzo-Bravo, M.J., Ng, C.K.L., et al. (2013). A unique Oct4 interface is crucial for reprogramming to pluripotency. Nat. Cell Biol. 15, 295–301.

44. Klemm, J.D., Rould, M.A., Aurora, R., Herr, W., and Pabo, C.O. (1994). Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. Cell 77, 21–32.

45. Chen, K., Long, Q., Xing, G., Wang, T., Wu, Y., Li, L., Qi, J., Zhou, Y., Ma, B., Schöler, H.R., et al. (2020). Heterochromatin loosening by the Oct4 linker region facilitates Klf4 binding and iPSC reprogramming. EMBO J. 39, e99165.

46. Han, D., Wu, G., Chen, R., Drexler, H.C.A., MacCarthy, C.M., Kim, K.P., Adachi, K., Gerovska, D., Mavrommatis, L., Bedzhov, I., et al. (2022). A balanced Oct4 interactome is crucial for maintaining pluripotency. Sci. Adv. 8, eabe4375.

47. Rand, T.A., Sutou, K., Tanabe, K., Jeong, D., Nomura, M., Kitaoka, F., Tomoda, E., Narita, M., Nakamura, M., Nakamura, M., et al. (2018). MYC releases early reprogrammed human cells from proliferation pause via retinoblastoma protein inhibition. Cell Rep. 23, 361–375.

48. Li, H., Ta, N., Long, C., Zhang, Q., Li, S., Liu, S., Yang, L., and Zuo, Y. (2019). The spatial binding model of the pioneer factor Oct4 with its target genes during cell reprogramming. Comput. Struct. Biotechnol. J. 17, 1226–1233.

49. Di Rocco, G., Gavalas, A., Popperl, H., Krumlauf, R., Mavilio, F., and Zappavigna, V. (2001). The recruitment of SOX/OCT complexes and the differential activity of HOXA1 and HOXB1 modulate the Hoxb1 auto-regulatory enhancer function. J. Biol. Chem. 276, 20506–20515.

50. Okumura-Nakanishi, S., Saito, M., Niwa, H., and Ishikawa, F. (2005). Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. J. Biol. Chem. 280, 5307–5317.

51. Ambrosetti, D.C., Basilico, C., and Dailey, L. (1997). Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. Mol. Cell. Biol. 17, 6321–6329.

52. 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. Mol. Cell. Biol. 19, 5453–5465.

53. Teif, V.B., Vainshtein, Y., Caudron-Herger, M., Mallm, J.P., Marth, C., Höfer, T., and Rippe, K. (2012). Genome-wide nucleosome positioning during embryonic stem cell development. Nat. Struct. Mol. Biol. 19, 1185–1192.

54. Dodonova, S.O., Zhu, F., Dienemann, C., Taipale, J., and Cramer, P. (2020). Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. Nature 580, 669–672.

55. Huertas, J., MacCarthy, C.M., Schöler, H.R., and Cojocaru, V. (2020). Nucleosomal DNA dynamics mediate Oct4 pioneer factor binding. Biophys. J. 118, 2280–2296.

56. Michael, A.K., Grand, R.S., Isbel, L., Cavadini, S., Kozicka, Z., Kempf, G., Bunker, R.D., Schenk, A.D., Graff-Meyer, A., Pathare, G.R., et al. (2020). Mechanisms of OCT4-SOX2 motif readout on nucleosomes. Science 368, 1460–1465.

57. MacCarthy, C.M., Huertas, J., Ortmeier, C., vom Bruch, H., Tan, D.S., Reinke, D., Sander, A., Bergbrede, T., Jauch, R., Schöler, H.R., et al. (2022). OCT4 interprets and enhances nucleosome flexibility. Nucleic Acids Res. 50, 10311–10327.

CellPress
OPEN ACCESS

Cell Stem Cell
Article

58. 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.e6.

59. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589.

60. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. 28, 495–501.

61. Chen, J., Zhang, Z., Li, L., Chen, B.C., 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.

62. Chen, J., Chen, X., Li, M., Liu, X., Gao, Y., Kou, X., Zhao, Y., Zheng, W., Zhang, X., Huo, Y., et al. (2016). Hierarchical Oct4 binding in concert with primed epigenetic rearrangements during somatic cell reprogramming. Cell Rep. 14, 1540–1554.

63. Von Hippel, P.H., Rees, W.A., Rippe, K., and Wilson, K.S. (1996). Specificity mechanisms in the control of transcription. Biophys. Chem. 59, 231–246.

64. Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., et al. (2012). A map of the cis-regulatory sequences in the mouse genome. Nature 488, 116–120.

65. Bentsen, M., Goymann, P., Schultheis, H., Klee, K., Petrova, A., Wiegandt, R., Fust, A., Preussner, J., Kuenne, C., Braun, T., et al. (2020). ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. Nat. Commun. 11, 4267.

66. Kang, L., Wang, J., Zhang, Y., Kou, Z., and Gao, S. (2009). iPS cells can support full-term development of tetraploid blastocyst-complemented embryos. Cell Stem Cell 5, 135–138.

67. Zhao, X.Y., Li, W., Lv, Z., Liu, L., Tong, M., Hai, T., Hao, J., Guo, C.L., Ma, Q.W., Wang, L., et al. (2009). iPS cells produce viable mice through tetraploid complementation. Nature 461, 86–90.

68. Nagy, A., Gócza, E., Diaz, E.M., Prideaux, V.R., Iványi, E., Markkula, M., and Rossant, J. (1990). Embryonic stem cells alone are able to support fetal development in the mouse. Development 110, 815–821.

69. Carey, B.W., Markoulaki, S., Hanna, J.H., Faddah, D.A., Buganim, Y., Kim, J., Ganz, K., Steine, E.J., Cassady, J.P., Creyghton, M.P., et al. (2011). Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. Cell Stem Cell 9, 588–598.

70. Buganim, Y., Markoulaki, S., van Wietmarschen, N., Hoke, H., Wu, T., Ganz, K., Akhtar-Zaidi, B., He, Y., Abraham, B.J., Porubsky, D., et al. (2014). The developmental potential of iPSCs is greatly influenced by reprogramming factor selection. Cell Stem Cell 15, 295–309.

71. Chen, J., Gao, Y., Huang, H., Xu, K., Chen, X., Jiang, Y., Li, H., Gao, S., Tao, Y., Wang, H., et al. (2015). The combination of Tet1 with Oct4 generates high-quality mouse-induced pluripotent stem cells. Stem Cells 33, 686–698.

72. Yoshioka, N., Gros, E., Li, H.R., Kumar, S., Deacon, D.C., Maron, C., Muotri, A.R., Chi, N.C., Fu, X.D., Yu, B.D., et al. (2013). Efficient generation of human iPSCs by a synthetic self-replicative RNA. Cell Stem Cell 13, 246–254.

73. Okita, K., Matsumura, Y., Sato, Y., Okada, A., Morizane, A., Okamoto, S., Hong, H., Nakagawa, M., Tanabe, K., Tezuka, K., et al. (2011). A more efficient method to generate integration-free human iPS cells. Nat. Methods 8, 409–412.

74. Han, L., Wei, X., Liu, C., Volpe, G., Zhuang, Z., Zou, X., Wang, Z., Pan, T., Yuan, Y., Zhang, X., et al. (2022). Cell transcriptomic atlas of the non-human primate Macaca fascicularis. Nature 604, 723–731.

75. Wunderlich, S., Kircher, M., Vieth, B., Haase, A., Merkert, S., Beier, J., Göhring, G., Glage, S., Schambach, A., Curnow, E.C., et al. (2014). Primate iPS cells as tools for evolutionary analyses. Stem Cell Res. 12, 622–629.

76. Wunderlich, S., Haase, A., Merkert, S., Beier, J., Schwanke, K., Schambach, A., Glage, S., Göhring, G., Curnow, E.C., and Martin, U. (2012). Induction of pluripotent stem cells from a cynomolgus monkey using a polycistronic simian immunodeficiency virus-based vector, differentiation toward functional cardiomyocytes, and generation of stably expressing reporter lines. Cell. Reprogram. 14, 471–484.

77. Kinoshita, M., Kobayashi, T., Planells, B., Klisch, D., Spindlow, D., Masaki, H., Bornelöv, S., Stirparo, G.G., Matsunari, H., Uchikura, A., et al. (2021). Pluripotent stem cells related to embryonic disc exhibit common self-renewal requirements in diverse livestock species. Development 148, dev199901.

78. Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. 33, 5868–5877.

79. Takikawa, S., Ray, C., Wang, X., Shamis, Y., Wu, T.Y., and Li, X. (2013). Genomic imprinting is variably lost during reprogramming of mouse iPS cells. Stem Cell Res. 11, 861–873.

80. Keshet, G., and Benvenisty, N. (2021). Large-scale analysis of imprinting in naive human pluripotent stem cells reveals recurrent aberrations and a potential link to FGF signaling. Stem Cell Rep. 16, 2520–2533.

81. Bar, S., Schachter, M., Eldar-Geva, T., and Benvenisty, N. (2017). Large-scale analysis of loss of imprinting in human pluripotent stem cells. Cell Rep. 19, 957–968.

82. Holm, T.M., Jackson-Grusby, L., Brambrink, T., Yamada, Y., Rideout, W.M., and Jaenisch, R. (2005). Global loss of imprinting leads to widespread tumorigenesis in adult mice. Cancer Cell 8, 275–285.

83. Jelinic, P., and Shaw, P. (2007). Loss of imprinting and cancer. J. Pathol. 211, 261–268.

84. Stadtfeld, M., Apostolou, E., Ferrari, F., Choi, J., Walsh, R.M., Chen, T., Ooi, S.S.K., Kim, S.Y., Bestor, T.H., Shioda, T., et al. (2012). Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells. Nat. Genet. 44, 398–405.

85. Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. Cell Stem Cell 4, 487–492.

86. Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D.G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature 448, 196–199.

87. Brons, I.G.M., Smithers, L.E., Trotter, M.W.B., Rugg-Gunn, P., Sun, B., Chuva De Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature 448, 191–195.

88. Choi, H.W.W., Joo, J.Y.Y., Hong, Y.J.J., Kim, J.S.S., Song, H., Lee, J.W.W., Wu, G., Schöler, H.R.R., and Do, J.T.T. (2016). Distinct enhancer activity of Oct4 in naive and primed mouse pluripotency. Stem Cell Rep. 7, 911–926.

89. Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hübner, K., and Schöler, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development 122, 881–894.

90. Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., et al. (2013). Derivation of novel human ground state naive pluripotent stem cells. Nature 504, 282–286.

91. Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., et al. (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell 15, 471–487.

Cell Stem Cell
Article

CellPress
OPEN ACCESS

92. Yang, S.H., Andrabi, M., Biss, R., Murtuza Baker, S., Iqbal, M., and Sharrocks, A.D. (2019). ZIC3 controls the transition from naive to primed pluripotency. Cell Rep. 27, 3215–3227.e6.

93. Adachi, K., Kopp, W., Wu, G., Heising, S., Greber, B., Stehling, M., Araúzo-Bravo, M.J., Boerno, S.T., Timmermann, B., Vingron, M., et al. (2018). Esrrb unlocks silenced enhancers for reprogramming to naive pluripotency. Cell Stem Cell 23, 266–275.e6.

94. Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development 136, 1063–1069.

95. Illich, D.J., Zhang, M., Ursu, A., Osorno, R., Kim, K.P., Yoon, J., Araúzo-Bravo, M.J., Wu, G., Esch, D., Sabour, D., et al. (2016). Distinct signaling requirements for the establishment of ESC pluripotency in late-stage EpiSCs. Cell Rep. 15, 787–800.

96. Bernemann, C., Greber, B., Ko, K., Sterneckert, J., Han, D.W., Araúzo-Bravo, M.J., and Schöler, H.R. (2011). Distinct developmental ground states of epiblast stem cell lines determine different pluripotency features. Stem Cells 29, 1496–1503.

97. Han, D.W., Greber, B., Wu, G., Tapia, N., Araúzo-Bravo, M.J., Ko, K., Bernemann, C., Stehling, M., and Schöler, H.R. (2011). Direct reprogramming of fibroblasts into epiblast stem cells. Nat. Cell Biol. 13, 66–71.

98. Han, D.W., Tapia, N., Joo, J.Y., Greber, B., Araúzo-Bravo, M.J., Bernemann, C., Ko, K., Wu, G., Stehling, M., Do, J.T., et al. (2010). Epiblast stem cell subpopulations represent mouse embryos of distinct pregastrulation stages. Cell 143, 617–627.

99. Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature 453, 519–523.

100. Han, D.W., Tapia, N., Araúzo-Bravo, M.J., Lim, K.T., Kim, K.P., Ko, K., Lee, H.T., and Schöler, H.R. (2013). Sox2 level is a determinant of cellular reprogramming potential. PLoS One 8, e67594.

101. Corsinotti, A., Wong, F.C., Tatar, T., Szczerbinska, I., Halbritter, F., Colby, D., Gogolok, S., Pantier, R., Liggat, K., Mirfazeli, E.S., et al. (2017). Distinct SoxB1 networks are required for naïve and primed pluripotency. eLife 6, e27746.

102. Kilens, S., Meistermann, Di, Moreno, Di. C.C., Gaignerie, A., Reignier, A., Lelièvre, Y., Casanova, M., Vallot, C., Nedellec, S., et al. (2018). Parallel derivation of isogenic human primed and naive induced pluripotent stem cells. Nat. Commun. 9, 1–13.

103. Lea, R.A., McCarthy, A., Boeing, S., Fallesen, T., Elder, K., Snell, P., Christie, L., Adkins, S., Shaikly, V., Taranissi, M., et al. (2021). KLF17 promotes human naïve pluripotency but is not required for its establishment. Development 148, dev199378.

104. Shahbazi, M.N., Scialdone, A., Skorupska, N., Weberling, A., Recher, G., Zhu, M., Jedrusik, A., Devito, L.G., Noli, L., Macaulay, I.C., et al. (2017). Pluripotent state transitions coordinate morphogenesis in mouse and human embryos. Nature 552, 239–243.

105. Guo, G., Von Meyenn, F., Santos, F., Chen, Y., Reik, W., Bertone, P., Smith, A., and Nichols, J. (2016). Naive pluripotent stem cells derived directly from isolated cells of the human inner cell mass. Stem Cell Rep. 6, 437–446.

106. Bredenkamp, N., Stirparo, G.G., Nichols, J., Smith, A., and Guo, G. (2019). The cell-surface marker sushi containing domain 2 facilitates establishment of human naive pluripotent stem cells. Stem Cell Rep. 12, 1212–1222.

107. Wojdyla, K., Collier, A.J., Fabian, C., Nisi, P.S., Biggins, L., Oxley, D., and Rugg-Gunn, P.J. (2020). Cell-surface proteomics identifies differences in signaling and adhesion protein expression between naive and primed human pluripotent stem cells. Stem Cell Rep. 14, 972–988.

108. Bi, Y., Tu, Z., Zhou, J., Zhu, X., Wang, H., Gao, S., and Wang, Y. (2022). Cell fate roadmap of human primed-to-naive transition reveals preimplantation cell lineage signatures. Nat. Commun. 13, 3147.

109. Zheng, C., Hu, Y., Sakurai, M., Pinzon-Arteaga, C.A., Li, J., Wei, Y., Okamura, D., Ravaux, B., Barlow, H.R., Yu, L., et al. (2021). Cell competition constitutes a barrier for interspecies chimerism. Nature 592, 272–276.

110. Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., et al. (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell 158, 1254–1269.

111. Liu, X., Nefzger, C.M., Rossello, F.J., Chen, J., Knaupp, A.S., Firas, J., Ford, E., Pflueger, J., Paynter, J.M., Chy, H.S., et al. (2017). Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming. Nat. Methods 14, 1055–1062.

112. Hanna, J., Cheng, A., and Saha, K. (2010). Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. Proc. Natl. Acad. Sci. USA 107, 9222–9227.

113. Yamauchi, K., Ikeda, T., Hosokawa, M., Nakatsuji, N., Kawase, E., Chuma, S., Hasegawa, K., and Suemori, H. (2020). Overexpression of nuclear Receptor 5A1 induces and maintains an intermediate state of conversion between primed and naive pluripotency. Stem Cell Rep. 14, 506–519.

114. Qin, H., Hejna, M., Liu, Y., Percharde, M., Wossidlo, M., Blouin, L., Durruthy-Durruthy, J., Wong, P., Qi, Z., Yu, J., et al. (2016). YAP induces human naive pluripotency. Cell Rep. 14, 2301–2312.

115. Yang, B.X.X., El Farran, C.A., Guo, H.C.C., Yu, T., Fang, H.T.T., Wang, H.F.F., Schlesinger, S., Seah, Y.F.S.F.S., Goh, G.Y.L.Y.L., Neo, S.P.P., et al. (2015). Systematic identification of factors for provirus silencing in embryonic stem cells. Cell 163, 230–245.

116. Eakin, G.S., and Hadjantonakis, A.K. (2006). Production of chimeras by aggregation of embryonic stem cells with diploid or tetraploid mouse embryos. Nat. Protoc. 1, 1145–1153.

117. Hiura, H., Toyoda, M., Okae, H., Sakurai, M., Miyauchi, N., Sato, A., Kiyokawa, N., Okita, H., Miyagawa, Y., Akutsu, H., et al. (2013). Stability of genomic imprinting in human induced pluripotent stem cells. BMC Genet. 14, 32.

118. Boland, M.J., Hazen, J.L., Nazor, K.L., Rodriguez, A.R., Gifford, W., Martin, G., Kupriyanov, S., and Baldwin, K.K. (2009). Adult mice generated from induced pluripotent stem cells. Nature 461, 91–94.

119. Hamanaka, S., Yamaguchi, T., Kobayashi, T., Kato-Itoh, M., Yamazaki, S., Sato, H., Umino, A., Wakiyama, Y., Arai, M., Sanbo, M., et al. (2011). Generation of germline-Competent rat induced pluripotent stem cells. PLoS One 6, e22008.

120. Lee, J., Go, Y., Kang, I., Han, Y.-M., and Kim, J. (2010). Oct-4 controls cell-cycle progression of embryonic stem cells. Biochem. J. 426, 171–181.

121. Hochedlinger, K., Yamada, Y., Beard, C., and Jaenisch, R. (2005). Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. Cell 121, 465–477.

122. Liu, K., Wang, F., Ye, X., Wang, L., Yang, J., Zhang, J., and Liu, L. (2014). KSR-based medium improves the generation of high-quality mouse iPS cells. PLoS One 9, e105309.

123. Buckberry, S., Liu, X., Poppe, D., Tan, J.P., Sun, G., Chen, J., Nguyen, T.V., de Mendoza, A., Pflueger, J., Frazer, T., et al. (2023). Transient naive reprogramming corrects hiPS cells functionally and epigenetically. Nature 620, 863–872.

124. Scognamiglio, R., Cabezas-Wallscheid, N., Thier, M.C., Altamura, S., Reyes, A., Prendergast, Á.M., Baumgärtner, D., Carnevalli, L.S., Atzberger, A., Haas, S., et al. (2016). Myc depletion induces a pluripotent dormant state mimicking diapause. Cell 164, 668–680.

125. Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature 292, 154–156.

126. Boroviak, T., Loos, R., Lombard, P., Okahara, J., Behr, R., Sasaki, E., Nichols, J., Smith, A., and Bertone, P. (2015). Lineage-specific profiling delineates the emergence and progression of naive pluripotency in mammalian embryogenesis. Dev. Cell 35, 366–382.

127. Thomas, K.R., and Capecchi, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell 51, 503–512.

128. Hall, B., Limaye, A., and Kulkarni, A.B. (2009). Overview: generation of gene knockout mice. Curr. Protoc. Cell Biol. *44*, 1–23.

129. Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S., and Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature *330*, 576–578.

130. Guo, G., von Meyenn, F., Rostovskaya, M., Clarke, J., Dietmann, S., Baker, D., Sahakyan, A., Myers, S., Bertone, P., Reik, W., et al. (2017). Epigenetic resetting of human pluripotency. Development *144*, 2748–2763.

131. Mazid, M.A., Ward, C., Luo, Z., Liu, C., Li, Y., Lai, Y., Wu, L., Li, J., Jia, W., Jiang, Y., et al. (2022). Rolling back of human pluripotent stem cells to an 8-cell embryo-like stage. Nature *605*, 315–324.

132. Gao, X., Nowak-Imialek, M., Chen, X., Chen, D., Herrmann, D., Ruan, D., Chen, A.C.H., Eckersley-Maslin, M.A., Ahmad, S., Lee, Y.L., et al. (2019). Establishment of porcine and human expanded potential stem cells. Nat. Cell Biol. *21*, 687–699.

133. Wu, J., Platero-Luengo, A., Sakurai, M., Sugawara, A., Gil, M.A., Yamauchi, T., Suzuki, K., Bogliotti, Y.S., Cuello, C., Morales Valencia, M., et al. (2017). Interspecies chimerism with mammalian pluripotent stem cells. Cell *168*, 473–486.e15.

134. Cao, J., Li, W., Li, J., Mazid, M.A., Li, C., Jiang, Y., Jia, W., Wu, L., Liao, Z., Sun, S., et al. (2023). Live birth of chimeric monkey with high contribution from embryonic stem cells. Cell *186*, 4996–5014.e24.

135. Alves-Lopes, J.P., Wong, F.C.K., Tang, W.W.C., Gruhn, W.H., Ramakrishna, N.B., Jowett, G.M., Jahnukainen, K., and Surani, M.A. (2023). Specification of human germ cell fate with enhanced progression capability supported by hindgut organoids. Cell Rep. *42*, 111907.

136. Choi, J., Huebner, A.J., Clement, K., Walsh, R.M., Savol, A., Lin, K., Gu, H., Di Stefano, B., Brumbaugh, J., Kim, S., et al. (2017). Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. Nature *1*, 1–7.

137. Dong, C., Beltcheva, M., Gontarz, P., Zhang, B., Popli, P., Fischer, L.A., Khan, S.A., Park, K.M., Yoon, E.J., Xing, X., et al. (2020). Derivation of trophoblast stem cells from naïve human pluripotent stem cells. eLife *9*, 1–26.

138. Guo, G., Stirparo, G.G., Strawbridge, S.E., Spindlow, D., Yang, J., Clarke, J., Dattani, A., Yanagida, A., Li, M.A., Myers, S., et al. (2021). Human naive epiblast cells possess unrestricted lineage potential. Cell Stem Cell *28*, 1040–1056.e6.

139. Tremble, K.C., Stirparo, G.G., Bates, L.E., Maskalenka, K., Stuart, H.T., Jones, K., Andersson-Rolf, A., Radzisheuskaya, A., Koo, B.K., Bertone, P., et al. (2021). Sox2 modulation increases naïve pluripotency plasticity. iScience *24*, 102153.

140. Sousa, J.A. De, Wong, C., Dunkel, I., Owens, T., Voigt, P., Hodgson, A., Baker, D., Schulz, E.G., Reik, W., Smith, A., et al. (2023). Epigenetic dynamics during capacitation of naïve human pluripotent stem cells. Sci Adv. *9*, eadg1936.

141. Kunitomi, A., Hirohata, R., Arreola, V., Osawa, M., Kato, T.M., Nomura, M., Kawaguchi, J., Hara, H., Kusano, K., Takashima, Y., et al. (2022). Improved Sendai viral system for reprogramming to naive pluripotency. Cell Rep. Methods *2*, 100317.

142. Fortunato, S., Adamski, M., Bergum, B., Guder, C., Jordal, S., Leininger, S., Zwafink, C., Rapp, H.T., and Adamska, M. (2012). Genome-wide analysis of the sox family in the calcareous sponge Sycon ciliatum: multiple genes with unique expression patterns. EvoDevo *3*, 14.

143. Siebert, S., Farrell, J.A., Cazet, J.F., Abeykoon, Y., Primack, A.S., Schnitzler, C.E., and Juliano, C.E. (2019). Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. Science *365*, 1–29.

144. Jager, M., Quéinnec, E., Le Guyader, H., and Manuel, M. (2011). Multiple Sox genes are expressed in stem cells or in differentiating neuro-sensory cells in the hydrozoan Clytia hemisphaerica. EvoDevo *2*, 12.

145. Bosch, T.C.G., and David, C.N. (1987). Stem cells of Hydra magnipapillata can differentiate into somatic cells and germ line cells. Dev. Biol. *121*, 182–191.

146. Sukparangsi, W., Morganti, E., Lowndes, M., Mayeur, H., Weisser, M., Hammachi, F., Peradziryi, H., Roske, F., Hölzenspies, J., Livigni, A., et al. (2022). Evolutionary origin of vertebrate OCT4/POU5 functions in supporting pluripotency. Nat. Commun. *13*, 5537.

147. Bakhmet, E.I., and Tomilin, A.N. (2022). The functional diversity of the POUV-class proteins across vertebrates. Open Biol. *12*, 220065.

148. Leichsenring, M., Maes, J., Mössner, R., Driever, W., and Onichtchouk, D. (2013). Pou5f1 transcription factor controls zygotic gene activation in vertebrates. Science *341*, 1005–1009.

149. Luo, L., Shi, Y., Wang, H., Wang, Z., Dang, Y., Li, S., Wang, S., and Zhang, K. (2022). Base editing in bovine embryos reveals a species-specific role of SOX2 in regulation of pluripotency. PLoS Genet. *18*, e1010307.

150. Liu, S., Bou, G., Sun, R., Guo, S., Xue, B., Wei, R., Cooney, A.J., and Liu, Z. (2015). Sox2 is the faithful marker for pluripotency in pig: evidence from embryonic studies. Dev. Dyn. *244*, 619–627.

151. Blassberg, R., Patel, H., Watson, T., Gouti, M., Metzis, V., Delás, M.J., and Briscoe, J. (2022). Sox2 levels regulate the chromatin occupancy of WNT mediators in epiblast progenitors responsible for vertebrate body formation. Nat. Cell Biol. *24*, 633–644.

152. Ormsbee Golden, B.D., Wuebben, E.L., and Rizzino, A. (2013). Sox2 expression is regulated by a negative feedback loop in embryonic stem cells that involves AKT signaling and FoxO1. PLoS One *8*, e76345.

153. Di, K.-Q., Gao, S., Cui, L.-F., Chang, G., Wu, F.-J., Ren, L.-K., An, L., Miao, K., Tan, K., Tao, L., et al. (2015). Generation of fully pluripotent female murine-induced pluripotent stem cells. Biol. Reprod. *92*, 123.

154. Arez M., Eckersley-Maslin M., Klobučar T., Gilsa Lopes J. von, Krueger F., Raposo A.C., Gendrel A.V., de Jesus B.B., and da Rocha S.T. (2020). Sex of donor cell and reprogramming conditions predict the extent and nature of imprinting defects in mouse iPSCs. Preprint at bioRxiv. https://www.biorxiv.org/content/10.1101/2020.11.20.370973v1

155. Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P., and Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for Sry. Nature *351*, 117–121.

156. Orzack, S.H., Stubblefield, J.W., Akmaev, V.R., Colls, P., Munné, S., Scholl, T., Steinsaltz, D., and Zuckerman, J.E. (2015). The human sex ratio from conception to birth. Proc. Natl. Acad. Sci. USA *112*, E2102–E2111.

157. Licciardi, F. (2021). In vitro fertilization gender predilection: more but Less. F. S Rep. *2*, 144.

158. Carrano, A., Juarez, J.J., Incontri, D., Ibarra, A., and Guerrero Cazares, H.G. (2021). Sex-specific differences in glioblastoma. Cells *10*, 1–22.

159. Veerapandian, V., Ackermann, J.O., Srivastava, Y., Malik, V., Weng, M., Yang, X., and Jauch, R. (2018). Directed evolution of reprogramming factors by cell selection and sequencing. Stem Cell Rep. *11*, 593–606.

160. Tan, D.S., Chen, Y., Gao, Y., Bednarz, A., Wei, Y., Malik, V., Ho, D.H.-H.H., Weng, M., Ho, S.Y., Srivastava, Y., et al. (2021). Directed evolution of an enhanced POU reprogramming factor for cell fate engineering. Mol. Biol. Evol. *38*, 2854–2868.

161. Bolton, H., Graham, S.J.L., Van der Aa, N., Kumar, P., Theunis, K., Fernandez Gallardo, E., Voet, T., and Zernicka-Goetz, M. (2016). Mouse model of chromosome mosaicism reveals lineage-specific depletion of aneuploid cells and normal developmental potential. Nat. Commun. *7*, 11165.

162. Liu, Z., and Kraus, W.L. (2017). Catalytic-independent functions of PARP-1 determine Sox2 pioneer activity at intractable genomic loci. Mol. Cell *65*, 589–603.e9.

163. Nowak-Imialek, M., Kues, W.A., Petersen, B., Lucas-Hahn, A., Herrmann, D., Haridoss, S., Oropeza, M., Lemme, E., Schöler, H.R., Carnwath, J.W., et al. (2011). Oct4-enhanced green fluorescent protein transgenic pigs: a new large animal model for reprogramming studies. Stem Cells Dev. *20*, 1563–1575.

164. Wuensch, A., Habermann, F.A., Kurosaka, S., Klose, R., Zakhartchenko, V., Reichenbach, H.-D., Sinowatz, F., McLaughlin, K.J., and Wolf, E.

(2007). Quantitative monitoring of pluripotency gene activation after somatic cloning in cattle. Biol. Reprod. *76*, 983–991.

165. Shahbazi, E., Moradi, S., Nemati, S., Satarian, L., Basiri, M., Gourabi, H., Zare Mehrjardi, N., Günther, P., Lampert, A., Händler, K., et al. (2016). Conversion of human fibroblasts to stably self-renewing neural stem cells with a single zinc-finger transcription factor. Stem Cell Rep. *6*, 539–551.

166. Baharvand, H., Ashtiani, S.K., Taee, A., Massumi, M., Valojerdi, M.R., Yazdi, P.E., Moradi, S.Z., and Farrokhi, A. (2006). Generation of new human embryonic stem cell lines with diploid and triploid karyotypes. Dev. Growth Differ. *48*, 117–128.

167. Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145–1147.

168. Okita, K., Yamakawa, T., Matsumura, Y., Sato, Y., Amano, N., Watanabe, A., Goshima, N., and Yamanaka, S. (2013). An efficient nonviral method to generate integration-free human-induced pluripotent stem cells from cord blood and peripheral blood cells. Stem Cells *31*, 458–466.

169. Webb, B., and Sali, A. (2016). Comparative protein structure modeling using MODELLER. Curr. Protoc. Bioinform. *54*, 1–37.

170. Bieniossek, C., Richmond, T.J., and Berger, I. (2008). MultiBac: multigene baculovirus-based eukaryotic protein complex production. Curr. Protoc. Protein Sci. *Chapter*, Unit 5.20.

171. Naviaux, R.K., Costanzi, E., Haas, M., and Verma, I.M. (1996). The pCL vector system: rapid production of helper-free, high-titer, recombinant retroviruses. J. Virol. *70*, 5701–5705.

172. Kime, C., Rand, T.A., Ivey, K.N., Srivastava, D., Yamanaka, S., and Tomoda, K. (2015). Practical integration-free episomal methods for generating human induced pluripotent stem cells. Curr. Protoc. Hum. Genet. *87*, 21.2.1–21.2.21.

173. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W., and Roder, J.C. (1993). Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. Proc. Natl. Acad. Sci. USA *90*, 8424–8428.

174. Hogan, B., Costantini, F., and Lacy, E. (1986). Manipulating the Mouse Embryo: a Laboratory Manual (Cold Spring Harbor Laboratory Press).

175. Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA *76*, 4350–4354.

176. Scholz, J., Besir, H., Strasser, C., and Suppmann, S. (2013). A new method to customize protein expression vectors for fast, efficient and background free parallel cloning. BMC Biotechnol. *13*, 12.

177. Lowary, P.T., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J. Mol. Biol. *276*, 19–42.

178. Klinker, H., Haas, C., Harrer, N., Becker, P.B., and Mueller-Planitz, F. (2014). Rapid purification of recombinant histones. PLoS One *9*, e104029.

179. Luger, K., Rechsteiner, T.J., and Richmond, T.J. (1999). Expression and purification of recombinant histones and nucleosome reconstitution. Methods Mol. Biol. *119*, 1–16.

180. Li, P., Song, L.F., and Merz, K.M. (2015). Systematic parameterization of monovalent ions employing the nonbonded model. J. Chem. Theor. Comput. *11*, 1645–1657.

181. Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J. Chem. Theor. Comput. *11*, 3696–3713.

182. Ivani, I., Dans, P.D., Noy, A., Pérez, A., Faustino, I., Hospital, A., Walther, J., Andrio, P., Goñi, R., Balaceanu, A., et al. (2016). Parmbsc1: A refined force field for DNA simulations. Nat. Methods *13*, 55–58.

183. Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. J. Comput. Chem. *26*, 1781–1802.

184. Court, F., Tayama, C., Romanelli, V., Martin-Trujillo, A., Iglesias-Platas, I., Okamura, K., Sugahara, N., Simón, C., Moore, H., Harness, J.V., et al. (2014). Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. Genome Res. *24*, 554–569.

185. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

186. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, 1–4.

187. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.

188. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

189. Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE blacklist: identification of problematic regions of the genome. Sci. Rep. *91*, 1–5.

190. Castro-Mondragon, J.A., Riudavets-Puig, R., Rauluseviciute, I., Lemma, R.B., Turchi, L., Blanc-Mathieu, R., Lucas, J., Boddie, P., Khan, A., Manosalva Pérez, N., et al. (2022). JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *50*, D165–D173.

191. Rosety, I., Zagare, A., Saraiva, C., Nickels, S., Antony, P., Almeida, C., Glaab, E., Halder, R., Velychko, S., Rauen, T., et al. (2023). Impaired neuron differentiation in GBA-associated Parkinson's disease is linked to cell cycle defects in organoids. npj Parkinsons Dis *9*, 166. https://www.nature.com/articles/s41531-023-00616-8.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Goat anti-Oct4	Santa Cruz Biotechnology	Cat# sc-8628; RRID: AB_653551
Mouse anit-Oct4	BD Biosciences	Cat# 611203; RRID: AB_398737
Rabbit anti-mouse-Oct4	Cell Signaling Tech	Cat# 83932; RRID: AB_2721046
Goat anti-Sox2	Santa Cruz Biotechnology	Cat# sc-17320; RRID: AB_2286684
Goat anti-Sox2	Neuromics	Cat# GT15098; RRID: AB_1623028
Rabbit anti-KLF17	Atlas Antibodies	HPA024629 RRID: AB_1668927
Mouse anti-human-SUSD2	Biolegend	Cat# 327401; RRID: AB_940656
Mouse anti-human-SUSD2	Miltenyi Biotec	Cat# 130-127-902; RRID: AB_2905407
Chemicals, oligos, and recombinant proteins		
PD0325901	Cayman Chemical	13034
XAV939	Sigma	X3004
Human LIF	In-house or Peprotech	N/A or 300-05
RSeT naïve-like media	STEMCELL Tech	05978
StemFlex primed media	Gibco	A3349401
Nucleofector kit	Lonza	VPH-5012 or V4LP-3002
Lipofectamine™ Stem Reagent	Invitrogen	STEM00001
FuGENE6	Promega	E269A
RT-qPCR oligos	Sigma	Table S2
EMSA DNA sequences	Sigma	Table S3
Critical commercial assays		
Zymo-Seq RRBS Library Kit	Zymo Research	D5461
SYBRGreen qPCR kit	Bio-Rad	1725125
Experimental models: Cell lines		
HEK293T cells	ATCC	CRL-3216
Rosa25rtTA-Gof18 MEFs	Derived in house	N/A
Cynomolgus fibroblasts	Wunderlich et al.[77]	N/A
Porcine embryonic fibroblasts	Nowak-Imialek et al.[164]	N/A
Bovine embryonic fibroblasts	Wuensch et al.[165]	N/A
Human foreskin fibroblasts	Shahbazi et al.[166]	N/A
Human 56 yo dermal fibroblasts	Coriell	AG04148
Human episomal iPSC line	Gibco	A18945
Human ESCs (H6)	Baharvand et al.[167]	N/A
Human ESCs (H9)	Thomson et al.[168]	N/A
Mouse E3 EpiSCs	Han et al.[99]	N/A
Experimental models: Organisms/strains		
Stbl2™ competent E. coli	Invitrogen	10268019
NEB® Stable competent E. coli	NEB	C3040
High Five™ Cells	Gibco	B85502
Sf9 cells	Gibco	11496015
C57BL/6×C3H mice	Bred in house	N/A
CD1 mice	Bred in house	N/A
SCID mice	Bred in house	N/A
Recombinant DNA		
Lentiviral pHAGE2-tetO-SKM/OSKM	Velychko et al.[13]	Addgene #136541, 136551

(Continued on next page)

Cell Stem Cell
Article

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Episomal pCXWB-EBNA1	Okita et al.[169]	Addgene #37624
Episomal pCXLE-mp53DD	Okita et al.[169]	Addgene #41859
Episomal pCXLE-ML	Okita et al.[74]	Addgene #27080
Episomal pCXLE-OCT4	Okita et al.[74]	Addgene #27076-27077
Episomal pCXLE	This study	Addgene #193290-193298
Lentiviral pHAGE2-tetO	This study	Addgene # 193345-49, 193299
Self-replicating RNA T7-VEE	This study	Addgene #193355-193360
Retroviral pMX	This study	Addgene #193350-193354
Deposited data		
MDS	This study	datashare.mpcdf.mpg.de/s/hovecsrp8aaONTR
ChIP-seq	This study	GEO: GSE247048
RNA-seq	This study	GEO: GSE247049
RRBS	This study	GEO: GSE247050
ATAC-seq of time-course priming	Yang et al.[93]	ArrayExpress: E-MTAB-7207
ATAC-seq of MEFs vs. ESCs	Li et al.[65]	GEO: GSE93029
Software and algorithms		
GREAT	McLean et al.[60]	great.stanford.edu
TOBIAS footprinting analysis	Bentsen et al.[65]	github.com/loosolab/TOBIAS
MODELLER	Webb and Sali[170]	salilab.org/modeller/

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Sergiy Velychko (Sergiy_Velychko@hms.harvard.edu).

Materials availability
Plasmids generated in this study have been deposited to Addgene (#193290-210020).

Data and code availability
- ChIP-seq, RNA-seq, RRBS data have been deposited at GEO (GSE247051) and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. The DOI is listed in the key resources table.
- This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice
All mice used were bred and housed at the mouse facility of the Max Planck Institute in Münster. Animal handling was in accordance with MPI animal protection guidelines.

The surrogate mouse embryos for tetraploid complementations were obtained by breeding super-ovulated B6C3 F1 females with CD1 males, a pairing that results in yellow coat color, and the surrogate mothers were pseudopregnant CD1 females (white). Rosa26-rtTA/Gof18 miPSCs have dark brown coats. While we present tetraploid complementation data for both sexes, for direct comparison between OSKM versus OS^{AV}KM cocktails we focused on male iPSCs, as male PSC lines have higher developmental potential. The experiment on the female mESC line (C57BL/6J background) illustrates that our findings apply for both sexes.

Primary cells
Mouse embryonic fibroblasts (MEFs, OG2 or Rosa26rtTA-Gof18, mixed sexes), human (newborn foreskin fibroblasts[166] and 56-year-old male dermal fibroblasts, Coriell, AG04148), cynomolgus macaque (female, MHH Hannover),[77] and porcine fetal fibroblasts

(male[164]) were cultured in high-glucose DMEM (Sigma) supplemented with 15% FBS, 1% Glutamax, 1% penicillin-streptomycin, 1% nonessential amino acids (NEAA), 1% sodium pyruvate (Sigma), and 1% β-mercaptoethanol (Gibco); bovine fetal fibroblasts (GOF 451-1, male)[165] were cultured in 50:50 DMEM/F12 (Gibco) and IMDM (with HEPES, Cytiva) with 15% FBS and the same supplements; 5 ng/ml of human bFGF (Peprotech) was used to improve cynomolgus, bovine, and porcine fibroblast cultures.

Cell lines

HEK293T cells were cultured in low-glucose DMEM (Sigma) supplemented with 10% FBS (Capricorn Scientific, ESC tested), 1% Glutamax, 1% penicillin-streptomycin, and 1% nonessential amino acids (all from Sigma).

Mouse naïve pluripotent stem cells were grown in KSR-based mouse embryonic stem cell (mESC) media: high-glucose DMEM medium supplemented with 15% KSR (Invitrogen), 1% Glutamax, 1% NEAA, 1% penicillin-streptomycin, 1% β-mercaptoethanol, and 20 ng/ml human recombinant LIF (purified in-house) on Mitomycin C-inactivated C3H MEF feeder layer. For 4N-complementation experiments the KSR-LIF media was supplemented with 2i (1 μM PD0325901 and 3 μM CHIR99021) for one passage. Mouse Gof18 Oct4-GFP reporter (E3) epiblast stem cells (EpiSCs)[99] were cultured in StemFlex™ media (Gibco) on FBS-coated dishes.

Human PSCs (H6[167] and H9[168] ESCs, female episomal A18945 iPSC line from Gibco, and iPSC derived in this study) were cultured in human ESC (hESC) media: either in DMEM/F12 supplemented with 15% KSR, 1% Glutamax, 1% NEAA, 1% penicillin-streptomycin, 1% β-mercaptoethanol, and 5 ng/ml bFGF or in StemFlex media (Gibco) on Matrigel-coated dishes (Corning), on FBS-coated dishes, or on Mitomycin C-inactivated CF1 MEF feeder layer. Mouse primed EpiSCs (E3[99]) were cultured in in StemFlex media. Cynomolgus iPSCs (ciPSCs) were derived and cultured in StemFlex media on Mitomycin C-inactivated CF1 MEF feeder layer. Bovine and porcine iPSCs (biPSCs and piPSCs) were derived and cultured in StemFlex media (Gibco) supplemented with 2 μM XAV939 (Sigma) on Mitomycin C-inactivated CF1 MEF feeder layer (prepared in house) in a hypoxic 5% O$_2$, 5% CO$_2$ incubator at 37°C; other cells were cultured in normoxic conditions. They were split on FBS-coated dishes with no feeders for karyotyping.

For naïve reset hiPSCs, the media was changed to RSeT™ (STEMCELL Technologies) at day 2, or StemFlex was kept (as indicated). For naïve reset biPSCs, the media was changed to mESCs supplemented with 1 μM PD0325901 (Cayman Chemical) and 2 μM XAV939 (Sigma).

Pluripotent stem cells of all five species were passed using Accutase (Sigma). 10 μM Rho-associated kinase inhibitor (ROCKi, Y-27632, Abcam) was added for the first 24h after passaging of primed PSCs of all five species (extended to 48h for mouse EpiSCs). The cells were routinely tested for Mycoplasma contamination and tested negative.

High Five™ and Sf9 insect cells were grown in serum-free EX-CELL® 420 medium containing L-glutamine (Sigma) and maintained in suspension culture at 0.5-1x10^6 cells/mL. Cultures were incubated at 26°C shaking at 100-120 rpm depending on flask size in a refrigerated shaking incubator (AutoQ Biosciences – AQ-2402D).

Microbe strains

TOP10 chemically-competent *E. coli* grown in Luria broth (LB) was used for plasmid amplification. For baculovirus plasmid DNA amplification, DH10EMBacY[171] (a gift from Dr. Imre Berger) were plated on agar plates containing LOC media supplemented with 50 μg/mL kanamycin, 10 μg/mL tetracycline 7 μg/mL of gentamicin, Bluo-Gal 100 μg/mL, and 1 mM IPTG. Selected colonies were grown in LOC media supplemented with 50 μg/mL kanamycin, 10 μg/mL tetracycline, and 7 μg/mL of gentamicin (Sigma). Stbl2 (Invitrogen) or NEB Stable competent E. coli grown in LB supplemented with 100 μg/ml of ampicillin or carbenicillin were used for preparing episomal plasmids.

METHOD DETAILS

Vector construction

The pMX-Sox2/Sox17 chimeric TF vectors were based on Addgene #13367[1] and the tet-inducible pHAGE2-tetO-Oct4-P2A-Sox2-17-T2A-Klf4-E2A-cMyc (OS*KM) and pHAGE-tetO-Sox2-17-T2A-Klf4-E2A-cMyc (S*KM) vectors were based on Addgene #136551 and 136541, respectively.[13]

The self-replicating RNA vector T7-VEE-OKS*iG was based on Addgene #58974.[72] The mouse episomal vectors pCXLE-Oct4-P2A-Klf4-IRES-Sox2 (OKS) and pCXLE-Oct4-P2A-Klf4-IRES-Sox2-17 (OKS*), and human episomal vectors pCXLE-SOX2-P2A-KLF4 (SK), pCXLE-SOX2^{AV}-P2A-KLF4 (S^{AV}K), pCXLE-SOX2-17-P2A-KLF4 (S*K), pCXLE-mCherry-E2A-SOX2-P2A-KLF4 were based on Addgene #27078,[169] except the inefficient self-cleaving peptide F2A was replaced with P2A to avoid protein fusion. pCXLE plasmids showed much better yields when grown in Stbl2™ (Invitrogen) or NEB Stable competent E. coli.

The mouse and human protein sequences of Sox2^{AV} and Sox2-17 were (HMG-box domains are uppercase, Sox17 parts are in bold):

>Mouse Sox2^{A61V}

mynmmetelkppgpqqasgggggggnataaatggnqknspDRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRP FIDEAKRLR**V**LHMKEHPDYKYRPRRKTKTLMKKDKytlpggllapggnsmasgvgvgaglgagvnqrmdsyahmngwsngsysmmqeqlgypqhpglnah gaaqmqpmhrydvsalqynsmtssqtymngsptysmsysysqqgtpgmalgsmgsvvkseasssppvvtssshsrapcqagdlrdmismylpgaevpepaapsrlhmaqh yqsgpvpgtaingtlplshm

>Human SOX2^A61V

mynmmetelkppgpqqtsgggggnstaaaaggnqknspDRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRPFI
DEAKRLR**V**LHMKEHPDYKYRPRRKTKTLMKKDKytlpggllapggnsmasgvgvgaglgagvnqrmdsyahmngwsngsysmmqdqlgypqhpglnahg
aaqmqpmhrydvsalqynsmtssqtymngsptysmsysqqgtpgmalgsmgsvvkseassspppvvtssshsrapcqagdlrdmismylpgaevpepaapsrlhmsqhy
qsgpvpgtaingtlplshm

>Mouse Sox2-17

mynmmetelkppgpqqasgggggggnataaatggnqknspDRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWK**ALTLA**EKR
PF**I**DEAKRLR**V**LHM**QD**HP**N**YKYRPRR**RKQVKRMKRV**eggflhalvepqagalgpeggrvamdglglpfpepgypagpplmsphmgphyrdcqglg
apaldgyplptpdtspldgveqdpaffaaplpgdcpaagtytyapvsdyavsveppagpmrvgpdpsgpampgilappsalhlyygamgspaasagrgfhaqpqq
plqpqappppppqqqhpahgpgqpspppealpcrdgtesnqptellgevdrtefeqylpfvykpemglpyqghdcgvnlsdshgaissvvsdassavyycnypdi

>Human SOX2-17

mynmmetelkppgpqqtsgggggnstaaaaggnqknspDRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWK**AL TLA**EKRPF
IDEAKRLR**V**LHM**QD**HP**N**YKYRPRR**RKQVKRLLKRV**eggflhglaepqaaalgpeggrvamdglglqfpeqgfpagpplLpphmgghyrdcqslgappld
gyplptpdtspldgvdpdpaffaapmpgdcpaagtysyaqvsdyagppeppagpmhprlgpepagpsipgllappsalhvyygamgspgagggrgfqmqpqh
qhqhqhqhhppgpgqpspppealpcrdgtdpsqpaellgevdrtefeqylhfvckpemglpyqghdsgvnlpdshgaissvvsdassavyycnypdv

iPSC generation and characterization

Mouse reprogramming experiments were done as described before.[13,36] Briefly, for retrovirus production monocistronic pMX-Oct4, Sox (Addgene #193350-193354), and Klf4 vectors were co-transfected with pCL-Eco (Addgene #12371)[172] in HEK293 cells with FuGENE6 (Promega) using low volume transfection protocol (Steffen et al., 2017). For lentivirus production, pHAGE2-tetO vectors were co-transfected with PAX2 and VSV. The viral supernatants were harvested after two and three days, filtered (Millex-HV 0.45 μm; Millipore) aliquoted and stored at -80°C. For reprogramming, Oct4-GFP MEFs (OG2 or Rosa26rtTA-Gof18) were plated on gelatin-coated 12-well plates at 3×10^4 cells per well in fibroblast media. A few hours later the cells were infected with titer-adjusted volumes of each viral supernatant supplemented with 6 μg/ml (final concentration) of protamine sulfate (Sigma). After two days, the media was replaced with mESC media. For mouse tet-inducible reprogramming, the cells were treated with Dox for 10 days (same as[13]), unless otherwise stated. Because all the reprogramming experiments were treated equally, enhanced kinetics of OS*KM reprogramming resulted in mature tetO-OS*KM expressing the Myc-containing transgene for much longer compared to tetO-OSKM or tetO-OS^AVKM iPSCs that emerged later in the 10-day course. This likely explains the poor quality of tetO-OS*KM versus tetO-OS^AVKM iPSCs. The 24h-iPSCs were derived from MEFs that by infecting them with lentiviral tetO-OS*KM and exposing them to Dox for 24h. Clonal iPSC colonies were picked after day 10, and propagated in the same manner as for other iPSC lines. We do not claim that reprogramming of MEF to iPSC was completed in just 24h. Rather, we posit that a 24-hour induction with OS*KM is sufficient to induce complete pluripotency.

For human retroviral reprogramming, 48h after infection, the transduced cells were split on a CF1 feeder layer at 10^4 per 6-well plate. After one week, fibroblast media was changed to hESC media.

For mouse episomal reprogramming 10^5 of Oct4-GFP (Rosa26TA-Gof18) MEFs were plated on gelatin-covered 6-well plates overnight and transfected with 1.5 μg of pCXLE-OKS or OKS* combined with 0.5 μg of pCXWB-EBNA1 (Addgene #37624) with FuGENE6.

Human self-replicating RNA-based reprogramming was performed as previously described.[72] Briefly, the T7-VEE constructs (Addgene #58974, 193356) were digested with *MluI* and then *in vitro* transcribed using RiboMAX Large Scale RNA Production System Kit (Promega). The transcripts were 2′-O-methylated, capped, and poly(A)-tailed using respective CELLSCRIPT kits following the manufacturer's protocol. For reprogramming, 1 μg of RNA replicons were transfected into 10^5 fibroblasts on 6-well plates using RiboJuice (Sigma) in the presence of 100 ng/ml B18R (Promega). The media was supplemented with 0.5 mM VPA, 5 μM EPZ to enhance the very inefficient RNA-based reprogramming. The reprogramming worked more efficiently when no puromycin selection was used. After two weeks, the cells were sorted for TRA-1-60 and plated on a CF1 feeder layer in hESC media without B18R (Figure 5I).

Human and cynomolgus macaque episomal reprogramming was done as previously described.[173] Briefly, 5×10^5 human newborn foreskin fibroblasts (Young, Y),[166] 56-year-old male dermal fibroblasts (Old, O, Coriell, AG04148), or macaque fibroblasts (MHH Hannover) were nucleofected with 3 μg of plasmid DNA mix: pCXLE-SOX2-P2A-KLF4 (Addgene #193292) or pCXLE-SOX2-17-P2A-KLF4 (Addgene #193290), pCXLE-L-MYC-F2A-LIN28 (ML, Addgene #27080), pCXLE-hOCT4-shTP53 (Addgene #27077), pCXWB-EBNA1 (Addgene #37624) using Lonza NHDF Nucleofector kit (U-23 program), and plated in ROCKi-containing fibroblast media on a CF1 feeder layer at different densities.

For livestock reprogramming, 10^6 bovine fetal fibroblasts (GOF 451-1)[165] or porcine fetal fibroblasts[164] were nucleofected with 6 μg of plasmid DNA mix: pCXLE-SOX2-P2A-KLF4 or pCXLE-SOX2-17-P2A, pCXLE-L-MYC-F2A-LIN28, pCXLE-hOCT4 (Addgene #27076), pCXWB-EBNA1 with or without p53DD (Addgene #41859) and plated on CF1 feeders.

The virus supernatant volumes were adjusted according to RT-qPCR titration using common WPRE or 3'UTR primers normalized to Rpl37a.[13] All the tetO lines were screened for promoter leaking, only those with minimal leaking were selected for characterization. The newly generated iPSC lines (mouse, human, cynomolgus, and cow) were karyotyped using DAPI staining of metaphase spreads, only the lines with correct chromosomal numbers were selected for characterization. As we reported before,[13] no difference in aneuploidy occurrence was observed between different cocktails. Similar to other studies, we only tested the quality of male iPSCs for this work.

Tetraploid (4N) complementation assay
Preparation of tetraploid embryos
Super ovulated B6C3 F1 females were mated with CD1 males. E1.5 embryos at the two-cell stage are flushed from the oviducts and collected in M2 medium.

After equilibration in fusion solution (0.3 M D-mannitol, 50 μM CaCl2, 0.3% BSA (Sigma)), 50-75 embryos are placed between the electrodes of a 250 μm gap electrode chamber (BLS Ltd.) containing 0.3 M mannitol with 0.3% BSA and fused with a Cellfusion CF-150/B apparatus (BLS Ltd.) with 0.5 mm Microslide (BTX-450). An initial electrical field of 2V is applied to the embryos followed by one peak pulses of 60V for 50 μs. Embryos are transferred back into KSOM-aa medium and immediately into a 37°C incubator with 5% CO_2. Embryos are observed for fusion after 15 to 60 minutes. The fused tetraploid embryos are cultured for 24h to the 4-cell stage under the same conditions.

Aggregation of iPSCs with zona-free embryos
- Preparation of aggregation plates for mouse embryos chimera production 1h before aggregation:

Using a KSOM medium filled 100μl-pipette, make 4 rows of microdrops (roughly 3mm in diameter) in a 35mm dish (Falcon, Cat. No. 35-3001), two drops in the first and fourth, five drops in the second and third rows.

Cover the whole plate with paraffin oil.
Sterilize the aggregation needle (BLS Ltd.) with 70% ethanol.
Press the aggregation needle into the plastic through the paraffin oil and culture medium, while making a circle movement to create a tiny scoop of about 300 μm in diameter with a clear smooth wall. Six to ten holes can be made within each droplet.

- iPSCs are aggregated and cultured with denuded 4-cell stage mouse tetraploid embryos as reported with a slight modification:[174]

Clumps of loosely connected iPSCs (15-20 cells in each) from short trypsin-treated day two iPSC cultures were chosen and transferred into microdrops of KSOM medium under mineral oil; each clump is placed in a depression in the microdrop. Meanwhile, batches of 30-50 embryos were briefly incubated in acidified Tyrode's solution[175] until dissolution of their zona pellucida. Two embryos were place on the iPSC clump. All aggregates are assembled in this manner, and cultured overnight at 37°C, 5% CO_2.

After 24h of culture, the majority of aggregates have formed blastocysts. Ten to fourteen embryos were transferred into one uterine horn of each 2.5 days post coitum, pseudopregnant CD1 female that had been mated with vasectomized males. For Cesarean Section, recipient mothers were sacrificed at E19.5 and pups were quickly removed. Newborns that were alive and respiring were cross fostered to lactating females.

Lentiviral naïve reset of human iPSCs
For primed-to-naïve reset (pluripotency upgrade), human iPSCs were transduced with monocistronic or polycistronic pHAGE2-EF1α lentiviral vectors carrying different subsets of Yamanaka factors from.[13] After two days, the cells were passed at low density (10^3 cells per 24-well plate) on an inactivated C3H feeder layer in mESC media supplemented with ROCKi with or without small molecules. 24h later the media was changed to mESC media (KSR-LIF) with ROCKi with or without 2i. Six days after passing, the cells were fixed and stained for KLF17 (HPA024629, ATLAS, 1:500). SK was the minimal subset that gave rise to KLF17$^+$ colonies, while neither Sox2 nor Klf4 alone did not.

Integration-free naïve reset of mouse epiblast stem cells
For integration-free reset, $3x10^5$ of GFP-negative Gof18 E3 mouse epiblast stem cells (mEpiSC)[99] cells were seeded in StemFlex+ROCKi media on FBS-coated 12-well plates and simultaneously transfected with 2 μg of episomal pCXLE-mCherry, pCXLE-mCherry-T2A-SOX2-P2A-KLF4 or pCXLE-mCherry-T2A-SOX2-17-P2A-KLF4 (Addgene #193293, #193296, and #193294, respectively) using 4μL of Lipofectamine Stem Reagent (Invitrogen, STEM00001) according to manufacturer's instructions; after 48h the cells were sorted for mCherry and plated in mESC media +ROCKi on an inactivated C3H feeder layer at 10^4 per 12-well plate. ~30% of sorted cells survived; of those ~50% of SK/S^{AV}K/S*K-transfected colonies grew dome-shaped and were GFP+/mCherry$^-$ already on day 4 after passing. The GFP$^+$ colonies were picked and clonally expanded for further characterization.

Integration-free naïve reset of human PSCs using episomal vectors
For human iPSC (hiPSCs, episomal A18945 from Gibco) or ESC (hESCs, H9) naïve reset, $0.5x10^6$ primed cells were nucleofected with a mix of 5 μg of pCXLE-mCherry-T2A-SOX2-17-P2A-KLF4 (Addgene #193294) or pCXLE-mCherry control (Addgene #193293) and 1 μg pCXWB-EBNA1 (Addgene #37624); with Nucleofector 2b (program B-016) and Lonza Human Stem Cell Nucleofector™ Kit 1 (Catalog #: VPH-5012) or 4D-Nucleofector (program CM-113) and Lonza P3 Primary Cell Kit L (Catalog #: V4LP-3002) according to manufacturer's protocol.

The nucleofected cells were plated on a dense feeder layer in StemFlex media supplemented with ROCKi. On the second day, the media was changed to StemFlex, and on day 3 to human naïve-like media (RSeT™, STEMCELL Technologies); the cells were fed daily.

Alternatively, the episomal S*K reset could be performed using feeder-free primed human iPSC culture conditions in StemFlex or E8 media.

The episomal vectors could also be delivered using Lipofectamin Stem Transfection Reagent (Invitrogen, STEM00001): hiPSCs were plated in feeder-free conditions on 6 well plate (10^6 cells per well) and next day the media was changed to pure OPTIM-MEM media supplemented with ROCKi, and the cells were transfected with 3 μg of pCXLE-mCherry-T2A-SOX2-17-P2A-KLF4 and 1 μg pCXWB-EBNA1 mixed with 8 μl of lipofectamine according to manufacturer's protocol for 4 hours. Following 4-hour transfection, the media was changed to StemFlex supplemented with ROCKi overnight for recovery. On day 2, the cells were dissociated using accutase, and split on feeders in StemFlex media supplemented with ROCKi. On day 3 the media was changed to RSeT.

While hypoxic conditions (5% O_2, 5% CO_2, 37°C) were used to generate naïve human cells for this study starting from day 2 after nucleofection, we later found that normoxic condition are favorable for human naïve reset. N2B27 media supplemented with PD, XAV, and LIF (PXL media) performed best for S*K naïve reset. The PXL media formulation was 1:1 mix of Neurobasal medium (Gibco, 21103049) and Advanced DMEM/F12 (Gibco, 11320082) supplemented with N2 (Gibco, 17502048), B27 minus vitamin A (Gibco, 12587010), sodium pyruvate (Gibco, 11360070), non-essential amino acids (Gibco, 11140050), GlutaMAX (Gibco, 35050061), Penicillin-Streptomycin (Gibco, 15070063), 0.1 mM b-mercaptoethanol (Gibco, 31350010), 50 μg/ml L-ascorbic acid (Sigma, A8960), 0.2% Geltrex (Gibco, A1413301), 1 μM PD0325901 (Cayman Chemical, 13034), 2 μM XAV939 (Sigma, X3004), 20 ng/ml human LIF (made in house or Peprotech, 300-05).

Increasing the ratio of pCXLE to pCXWB-EBNA1 up to 1:1 (3+3 μg for 100ul nucleofection reaction) could increase the longevity of the episome improving the efficiency of naïve reset, but could also be toxic for sensitive PSC lines. While SK is the minimal cocktail capable of inducing human naïve pluripotency, polycistronic SKM episomes[13] could further improve the reset efficiency (Addgene #210016-210018).

Gene expression analysis was performed using CYBR Green qPCR as previously described.[13] the oligos for human naive reset the primers can be found in Table S2.

The following plasmids were used to constitutively label the A18945 hiPSC line for mouse/human chimera experiments: AAVS1-Pur-CAG-mCherry (Addgene #80946), gRNA_AAVS1-T2 (Addgene #41818), pX330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene #42230). pCXLE-SOX2-17-P2A-KLF4 (Addgene #193290) was used for naïve reset not to interfere with constitutive RFP expression. For assessing embryo development contribution, we aggregated the S*K-reset hiPSCs with mouse morulae at E2.5 as previously described.[117] After 2 days of culture (E4.5), the chimeric embryos were stained for human-specific SUSD2 (Biolegend 327401) and mouse-specific Oct4 (D6C8T, Cell Signaling). All the experiments were performed in accordance with ISSCR guidelines.

Mammalian cell overexpression and whole-cell lysate (WCL) generation

HEK293 cells cultured on 10cm dishes were transfected with 10 μg of pLVTHM or pHAGE2 vectors under the control of an EF1α promoter and containing the WT or mutant versions of Oct4 or Sox2 with Fugene6 (Promega) using a low volume protocol (Steffen et al., 2017). Three days after transfection, the cells were dissociated from the plate using Accutase (Sigma), collected, counted, and washed with PBS. WCLs were generated by five cycles of freeze-thawing pellets resuspended in 12.5 μL per million cells in lysis buffer (20 mM HEPES-KOH pH 7.8, 150 mM NaCl, 0.2 mM EDTA pH 8, 25% glycerol, 1 mM DTT, and cOmplete™ protease inhibitor cocktail (Merck). After disruption, lysates were spun at 14k RCF at 4°C for 10 min. After centrifugation, pellets were discarded and the supernatants transferred to a new tube for further analysis. Protein concentrations were estimated by diluting samples in 0.1% SDS solution, measuring A_{230} and A_{260}, and applying the equation:

$$\text{Conc.}(\mu g / \mu L) = (0.183 * A_{230} - 0.075 * A_{260}) * \text{dilution factor}$$

All samples were diluted to 1 μg/μL, aliquoted, snap frozen, and stored at -80°C. Western blots were run to compare expression levels between mutants. Expression was evaluated by Quantity One® (v4.6.7, Bio-Rad) densitometry to adjust for equal amounts of expression using WCL of untransfected cells to maintain total protein content, when necessary.

Western blot analysis

5-10 μg of total protein was combined with Laemmli sample buffer, heated, and loaded onto 12% mini SDS-polyacrylamide gel (SDS-PAG) using the Towbin buffer system.[176] Gels were run initially at 15V for 15 minutes to load samples into the stacking gel and then 50V for 30-60 minutes to resolve the proteins of interest. Samples were transferred to Immobilin®-FL PVDF membranes (Merck Millipore Ltd.) at 4°C under 300V for 2h. Membranes were blocked for one hour at room temperature in 5% skim milk (Sigma) dissolved in PBS with 0.1% Tween-20 (PBS-T) and incubated overnight at 4°C with rotation in the primary antibody diluted in blocking solution. The following day the membrane was washed three times in PBS-T and then incubated in secondary antibody diluted in blocking solution for one hour at 25°C. The following antibodies were used: polyclonal goat anti-Oct4 N-19 (sc-8628, Santa Cruz Biotechnology) or monoclonal mouse anti-Oct4 (611203, BD Biosciences), polyclonal goat anti-Sox2 (sc-17320 from Santa Cruz Biotechnology or GT51098 from Neuromics), monoclonal mouse anti-alpha tubulin (T6199, Sigma), 647-conjugated anti-goat (Alexafluor), and 647-conjugated anti-mouse (Alexafluor). Western blot signal was detected using Fujifilm FLA-9000 fluorescence scanner (Fujifilm).

Insect cell expression and protein purification

The coding sequence of full-length *Mus musculus* Sox2 or Sox2^{AV} was cloned into pCoofy27 plasmid with an N-terminal 6xHis-tag using SLIC as previously described: forward primer 3C, reverse primer ccdB.[177] Plasmids were then transformed into DH10EMBacY (a gift from Dr. Imre Berger) for baculovirus plasmid DNA amplification.[171] Bacmids were purified using Macherey-Nagel Xtra BAC100

(Düren) and then transfected into a suspension of Sf9 cells at 0.8×10^6 cells/mL grown in serum-free EX-CELL® 420 medium containing L-glutamine (Sigma) and incubated at 26°C with shaking for virus production. Cells were monitored daily for increased cell size and GFP fluorescence. Once ~90% of cells were GFP+, viral suspensions were spun down and then filtered through 0.22 mm. Viral supernatants were expanded once before being used for infection, filtered aliquots were stored at -80°C.

Optimal protein expression conditions were determined empirically. Mid-log phase High Five™ insect cells were split to 10^6 cells/mL in 2 L and then infected with 10-12 mL of P1 baculovirus from previous steps per liter of cells. Following incubation at 28°C for 96h with shaking, cell pellets were collected by centrifugation. Pellets were resuspended in lysis buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 30 mM Imidazole, 5% glycerol, 0.1% Triton X-100, cOmplete™ protease inhibitor cocktail (Merck), and 1 mM DTT), frozen and thawed once, then sonicated at 4°C using a probe sonicator (Bandelin Sonopuls, Bandelin Eletronics). Pellets were resuspended in inclusion body wash buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 1 mM EDTA, 1% Triton X-100, cOmplete™ protease inhibitor cocktail (Merck), and 1 mM DTT) and subject to four cycles of Dounce homogenization followed by centrifugation for 20 min. at 18k RCF and 4°C, twice with inclusion body wash buffer and twice in buffer without Triton X-100. The final pellet was cut twice in DMSO and then incubated for 30 min at 25°C. Unfolding buffer (7 M guanidine hydrochloride, 20 mM Tris-HCl pH 7.5, 5 mM DTT) was added to the pellet and incubated while rotating for 1h at 25°C. Nickel Sepharose slurry (GE Healthcare) was washed and equilibrated in binding buffer, then supernatant was added and incubated at 4°C overnight with rotation. Proteins were eluted using the unfolding buffer with additional 500 mM imidazole. Eluate fractions were checked with SDS-PAGE and relevant fractions were pooled. Using 7 kDa molecular weight cut off (MWCO) dialysis tubing, pooled fractions were dialyzed for three buffer changes of at least 6 h for each volume of refolding buffer at 4°C (7 M urea, 20 mM Na Acetate pH 5.2, 200 mM NaCl, 1 mM EDTA, and 5 mM DTT). Following centrifugation to remove any insoluble material, the supernatant was dialyzed (7 kDa MWCO) in refolding buffer with decreasing amounts of urea: 1 h 6 M urea, 2h 4 M, 2h 2 M, and 1 h in size exclusion chromatography (SEC) buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 5% glycerol). Eluate was centrifuged to remove any precipitate before loading onto HiLoad 16/60 Superdex 200 SEC column (GE Healthcare).

The coding sequence for full-length Oct4 from *M. musculus* was cloned into the pOPIN expression vector using the SLIC method and Phusion Flash High-Fidelity PCR Master Mix (Finnzymes/New England Biolabs). SLIC reactions were then transformed into One Shot™ OmniMAC™ 2 T1® Chemically Competent *E. coli* (ThermoFisher Scientific). After sequencing, the pOPIN-cHis-Oct4 construct was co-transfected with flashBACULTRA™ bacmid DNA (Oxford Expression Technologies) into Sf9 cells (ThermoFisher Scientific) using Cellfectin II® (ThermoFisher Scientific) to generate recombinant baculovirus. Mid-log phase Sf9 cells were used to amplify the virus. Suspension High Five™ cells were infected with P3 virus for two days at 27°C and 120 rpm shaking. After expression, crude lysates were purified on a HiTrap TALON column (GE Healthcare), cleaved on the column with 3C protease followed by size exclusion chromatography (HiLoad Superdex 200, GE Healthcare). The final product was collected in 25 mM HEPES pH 7.8, 150 mM NaCl, 1 mM TCEP, and 5% glycerol with ~95% purity confirmed by SDS-PAGE. Fractions were checked with SDS-PAGE, pooled, and finally quantified using the NanoDrop spectrophotometer (ND-1000, ThermoFisher Scientific) and the Protein A_{280} program using specific molecular weight and extinction coefficients for either Sox2 or Oct4. Unless otherwise indicated all chemicals were from Sigma-Aldrich.

Electrophoretic mobility shift assays (EMSAs)

DNA probes were generated by annealing complementary 5' labeled Cy5 oligos (Metabion International AG) followed by purification from 10% polyacrylamide gels. EMSA DNA sequences can be found in Table S3. For binding reactions, WCL (2-4 ug of total protein) or purified proteins were incubated in binding buffer (25 mM HEPES-KOH pH 8, 50 mM NaCl, 0.5 mM EDTA, 0.07% Triton X-100, 4 mg/mL BSA, 7 mM DTT, and 10% glycerol) and 70 nM Cy5-dsDNA at 37°C for 1h. Samples were then loaded onto 6% native polyacrylamide gels (37.5/1 acrylamide/bis-acrylamide) containing 0.3x Tris-borate EDTA and 5% glycerol and run at 10 mA/gel in running buffer of the same composition. 5% native gels were used for the compressed motif experiments to resolve the Sox17 monomer from the lower non-specific band of the HEK293T cells.

The WCL EMSAs throughout the manuscript were generated from lysates of HEK293T cells overexpressing our proteins of interest. The system was optimized by screening several different cell lines, promoters, and transfection condition combinations to find the optimal overexpressed protein to background band ratio. The intensity of the background bands varied between transfections and the protein being overexpressed. All EMSAs were adjusted for equal amounts of the overexpressed proteins being compared based on western blotting and monomer binding. WCL of untransfected 293T cells was added to reactions to equalize the total protein in each lane.

Gels were imaged using Fujifilm FLA-9000 fluorescence scanner using (Fujifilm). Fraction bound was determined by densitometry of raw data using Quantity One® (v4.6.7, Bio-Rad) and the following equation for specific bands and then normalized: $F_B = DNA_{bound}/(DNA_{bound} + DNA_{unbound})$. Half-life was calculated using fraction bound as a function of protein concentration from at least two independent experiments, error bars represent SD.

For competition experiments, pre-formed protein/DNA or protein/nucleosome complexes (see binding conditions above) were loaded onto native gels (t=0) and then incubated with unlabeled double stranded DNA containing the *Nanog* locus. Protein dissociation was monitored by removing aliquots of the reaction at the given time points and loading them onto a running gel. Protein complex stability was highly variable thus conditions for competition assays were determined empirically and can be found in Table S4.

Supershift assays were run under the same conditions as equilibrium or static EMSAs, see above. After incubation of the proteins with DNA for 1 hour at 37°C, antibody was added to the reactions and incubated at room temperature for 30 min. Antibody/total

Cell Stem Cell
Article

protein ratios were empirically determined as 1 μg of antibody per 2 μg of total protein. For Oct4 and Sox2 detection, mouse anti-Oct4 from BD Biosciences (611203) and goat anti-Sox2 from Neuromics (GT51098) were used, respectively.

Nucleosome assembly
The nucleosome DNA sequence Widom +6 consists of 147 bp of the established Widom 601 sequence[178] with a Sox/Oct motif (**CTTTGTT**ATGCAAAT) at super helical location +6, with the nucleosome dyad being zero.[56] Double stranded DNA was purchased from IDT (Coralville) and labeled using Cy-5 conjugated primers via PCR, as previously described.[56] Nucleosomes were assembled in DNA:octamer ratios ranging from 1:1.2-1:1.6 with purified full-length *D. melanogaster* histone octamer[179] using the salt-gradient dialysis method previously described,[180] final buffer composition: 10 mM HEPES pH 7.6, 50 mM NaCl, 1 mM EDTA, and 0.5 mM DTT. Following dialysis, nucleosomes were heat shifted at 37°C for 2h. Nucleosome quality and concentration were evaluated using native PAGE run with a histone-free DNA standard curve made from the parent DNA. Histone stoichiometry was checked by 22% SDS-PAGE followed by Coomassie staining (R-250; SERVA). Nucleosome were stored at 4°C in the dark and used for no longer than three weeks from the date of assembly.

Molecular Dynamics Simulation (MDS)
We used the model of the Sox2/Oct4 heterodimer bound to a regulatory DNA element from the *Hoxb1* enhancer that we previously built[38] as template for building new models for the Sox2/Oct4, Sox2^{A61V}/Oct4, Sox2/Oct6, Sox2^{A61V}/Oct6. Using MODELLER[170]w (https://salilab.org/modeller/), we adapted the sequences, extended each model of the Oct factor with 4 and 8 residues at the N- and C-termini, respectively, and similarly with each model of the Sox factor with 4 and 5 residues. We built 100 models for each ternary complex using a "slow" optimization procedure that included a "slow" MD refinement as defined in MODELLER. We ranked the models using an energetic score ("DOPE") and selected 2 models for each complex for MD simulations. In each of these models, we extended the DNA by 16 and 18 base pairs at the 5' and 3' ends using the mouse *Hoxb1* sequence.[45] The final sequence in our model was:

5'-AGAGTGATTGAAGTGTCTTTGTCATGCTAATGATTGGGGGGGAGATGGAT-3'

Then, we solvated the systems in a truncated octahedron periodic box of SPCE water with the distance between any protein-DNA atom to the box edges larger than 12 Å. We added 73 neutralizing Na+ ions and 150 mM KCl (225 K+ and 225 Cl- ions). For the ions we used the parameters developed by Li and Merz.[181] We used the Amber-ff14SB[182] and the Amber-parmbsc1[183] force fields for proteins and DNA respectively. We energy minimized and equilibrated each system with a protocol described previously.[35] With each model we performed 2 independent, 1.2 μs long MD simulations by assigning different velocity distributions before the equilibration (in total 4 x 1.2 μs = 4.8 μs per system). We applied periodic boundary conditions in the isothermic-isobaric (NPT) ensemble with a timestep of 2 fs. The temperature was maintained at 300 K with Langevin Dynamics (damping coefficient of 0.1 ps-1). The pressure was maintained at 1 atm with the Nose Hoover Langevin Piston method with the period and decay of 1.2 and 1.0 ps, respectively. The direct calculation of the non-bonded interactions was truncated at 10 Å and the chemical bonds of hydrogens were kept rigid with the SHAKE algorithm. Long range electrostatics were calculated using the particle mesh Ewald algorithm. All simulations were performed in NAMD.[184] Snapshots were selected for analysis every 10 ps.

The coordination number between two atom selections describes the number contacts between the selections using a continuous switching function with a distance threshold for contact formation as implemented in the COLVAR module of NAMD. The mathematical formula is:

$$c = \sum_i \sum_j \frac{1 - (d_{ij}/d_0)^n}{1 - (d_{ij}/d_0)^m}$$

where i,j = a pair of atoms, one from each selection; d_{ij}= the distance between $i$and j; d_0= the distance threshold (4.5 Å); n,m= exponents describing the switching from contact to no contact ($n = 6, m = 12$).

To build the models of the Oct4-Sox2 complexes bound to the Nanog element, we started from the Oct4-Sox2-Hoxb1 model and adapted the sequence using the swapna function in Chimera (https://www.cgl.ucsf.edu/chimera/). The models were minimized and equilibrated with the same procedure. 2 independent simulations, each 2 microsecond long were performed with the same protocol (described above).

For the simulations of the Sox2/Oct4 and Sox2^{A61V}/Oct4 bound to the Pou5f1 distal enhancer (Oct4DE) we used the AlphaFold model of Oct4 instead of the model based on the structure of Oct4 bound to the PORE DNA element. AlphaFold predicts a different orientation of the linker α5 helix relative to the POU$_S$ domain. This orientation may represent more accurately the solution structure of Oct4 because in the PORE bound structure the linker orientation may be stabilized by the Oct4 dimerization. 3 starting models were built that differ in the conformation of the N-terminal and C-terminal extensions of the proteins. With each model we performed 3 independent simulations, two were 3 micros long and 1 was 2 micros long. The rest of the protocol was the same.

NGS and bioinformatic analysis
For ChIP-seq experiments, Rosa26TA-Gof18 MEFs were infected with titrated volumes of pHAGE2-tetO-Klf4-IRES-Sox2/Sox2^{AV} with or without pHAGE2-tetO-Oct4/Oct6. After 48h, the media was replaced with fibroblast media supplemented with doxycycline (dox). The samples were collected 48h after dox-induction. Gene ontology analysis was done by GREAT.[60]

RNA-seq and RRBS were performed for human iPSC at passage 10-12, human ESCs (H6[167] and H9[168]) at passage 35-36 grown in StemFlex media on Matrigel. The sample processing and data analysis for RNA-seq, ChIP-seq, RRBS were performed as described before.[13,22,81]

For RRBS analysis Fastq files were trimmed with Trim Galore (bioinformatics.babraham.ac.uk/projects/trim_galore/) flagging the options –RRBS. The trimmed fastq files were aligned to GRCh38 using bwameth (github.com/brentp/bwa-meth) and methylation metrics were extracted using MethylDackel (https://github.com/dpryan79/ MethylDackel), flagging the options –minDepth 10. The genomic coordinates of known imprinted DMRs[185] were converted to GRCh38 using the LiftOver tool from UCSC and the methylation levels of CpGs within these regions were extracted with bedtools intersect (bedtools.readthedocs.io/en/latest/content/tools/intersect.html). Finally, the mean methylation level for all CpGs of a given DMR was calculated. The several DMRs corresponding to SNRPN were collapsed as well. Only DMRs which were represented by all samples were taken into consideration for the comparative LOI analysis. A DMR was considered to lose imprinting if it showed less than 30% mean methylation levels.

For footprinting analysis, the publicly available data were aligned to mm10 genome using bowtie2[186] using "–very-sensitive -X 2000 –no-mixed" options; the mitochondrial and duplicate reads were removed, and the reads were sorted and indexed using samtools[187]; spearman correlation was plotted using deeptools[188]; the peaks were called using macs2 using "-g mm -f BAMPE –call-summits –cutoff-analysis –keep-dup all -B" options[189]; the output of macs2 was used for TOBIAS footprinting analysis[65] using ENCODE blacklist,[190] JASPAR MEME motif database[191] with some additional custom motifs.[22]

QUANTIFICATION AND STATISTICAL ANALYSIS

The statistical details of experiments can be found in the figure legends, figures, results, and method details. All EMSA experiments were performed at least three times. Protein complex half-lives were determined using nonlinear regression in GraphPad Prism 7 for Mac (version 7.0a) and only used when goodness of fit, evaluated by R^2 values, was 0.95 or greater. All reprogramming data are representative of at least three independent experiments, each with three biological replicates (n=3), data are shown as mean ± SD, unpaired two-tailed Student's t test was used to test the statistical significance of the data. Data were plotted and analyzed using GraphPad Prism 7. For ChIP-seq analysis p-values were calculated using the unpaired Wilcoxon rank sum test (R function pairwise.wilcox.test). For all statistical analysis: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$ and $****p < 0.0001$. Tetraploid complementation data were shown as mean between clonal iPSC lines generated with the same delivery and construct ±SEM, the mean of each line was plotted individually, raw numbers are available in Table S1.