

Published in final edited form as:

Nat Cell Biol. 2021 August 01; 23(8): 834–845. doi:10.1038/s41556-021-00727-5.

Dissecting OCT4 defines the role of nucleosome binding in pluripotency

Gareth A. Roberts^{#1}, Burak Ozkan^{#1}, Ivana Gachulinová¹, Michael R. O'Dwyer¹, Elisa Hall-Ponsole¹, Manoj Saxena², Philip J. Robinson², Abdenour Soufi^{1,*}

¹Centre for Regenerative Medicine, Institute for Regeneration and Repair, University of Edinburgh, 5 Little France Drive, Edinburgh, EH16 4UU, UK

²Institute of Structural and Molecular Biology, Department of Biological Sciences, Birkbeck, University of London, Malet Street, London, WC1E 7HX, UK

These authors contributed equally to this work.

Abstract

Pioneer transcription factors (TFs) such as OCT4 can target silent genes embedded in nucleosome-dense regions. How nucleosome interaction enables TFs to target chromatin and determine cell identity remains elusive. Here, we systematically dissect OCT4 to show that nucleosome binding is encoded within the DNA-binding domain and yet can be uncoupled from free DNA binding. Furthermore, accelerating the binding kinetics of OCT4 to DNA enhances nucleosome binding. In cells, uncoupling nucleosome binding diminishes the ability of OCT4 to individually access closed chromatin, while more dynamic nucleosome binding results in expansive genome scanning within closed chromatin. However, both uncoupling and enhancing nucleosome binding are detrimental to inducing pluripotency from differentiated cells. Remarkably, stable interactions between OCT4 and nucleosomes are continuously required for maintaining the accessibility of pluripotency enhancers in stem cells. Our findings reveal how the affinity and residence time of OCT4-nucleosome complexes modulate chromatin accessibility during cell fate changes and maintenance.

To maintain cell identity, TFs are often associated with accessible enhancers and promoters of active genes. However, during cell fate changes, certain TFs target silent genes within closed chromatin, acting as “pioneer factors”¹. For instance, Oct4, Sox2, Klf4 and c-Myc (OSKM) predominantly occupy open chromatin to maintain pluripotency in embryonic stem cells (ESCs)^{2–4}. Yet, during early reprogramming of fibroblasts to induced pluripotent stem

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: <https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms>

*Corresponding author and lead contact: Abdenour.Soufi@ed.ac.uk.

Author Contributions. A.S conceived the study and designed the experiments. G.A.R carried out recombinant protein purification, EMSA and nucleosome reconstitution with support from I.G. B.O carried out virus production and iPSCs reprogramming. G.A.R and B.O carried out ChIP-seq and ATAC-seq with support from M.R.O. E.H-P carried out molecular cloning, MEF generation, and pluripotency rescue with help from B.O and G.A.R. A.S, G.A.R and B.O carried out bioinformatic analysis with support from M.R.O. I.G. generated nucleosomes for cryo-EM. M.S and P.J.R performed cryo-EM data collection, LIN28B-nucleosome reconstruction, model building and refinement. A.S prepared the manuscript with contribution from G.A.R, B.O, E.H-P, M.R.O, and P.J.R.

Competing Interests. The authors declare no competing interests.

cells (iPSCs), OSK, but not c-Myc, act as pioneer factors⁵⁻⁹. Nonetheless, binding of OSK to open chromatin has also been suggested to be important for reprogramming^{10,11}. To date, it has not been possible to separate the conventional open chromatin binding from closed chromatin targeting, limiting the ability to directly examine the role of pioneer activity in pluripotency.

We have previously shown that OSK can directly interact with nucleosomes like the paradigm pioneer factor FoxA, supporting a link between pioneer activity and nucleosome binding¹²⁻¹⁷. A systematic evolution of ligands by exponential enrichment (SELEX) study has revealed that a wide variety of DNA-binding domains (DBDs) can bind nucleosomes *in vitro*¹⁸. Commonly, DBDs containing short anchoring α helices have been shown to interact most strongly with nucleosomes *in vitro*¹⁹. However, FoxA-DBD requires an extra helical region to bind nucleosomes and open chromatin, suggesting that DBDs may not contain the full pioneer capacity^{15,20}. Recently, the cryo-EM structure of OCT4-DBD co-bound with SOX2-DBD to an engineered nucleosome has shown a potential OCT4-SOX2 nucleosome readout that involves the OCT4 POU-specific domain (POU_S) and SOX2 high mobility group (HMG), but not the POU-homeodomain (POU_{HD}) of OCT4-DBD²¹. The cooperative interaction between OCT4 and SOX2 is critical for pluripotency in early development and during reprogramming²²⁻²⁷. Nevertheless, how OCT4 binds nucleosomes alone, which is the essential criteria for pioneer activity, remains elusive²⁸. Defining the pioneer function is critical to elucidating the mechanism by which TFs access closed chromatin.

In this study, we generated a mutant library spanning the full-length of human OCT4 to map nucleosome binding domains. We reveal that the interaction of OCT4 with nucleosomes plays a supportive rather than an obstructive role that facilitates and maintains chromatin accessibility during reprogramming and pluripotency maintenance.

Results

Systematic OCT4 mutagenesis maps nucleosome binding domains

OCT4 shows high amino-acid (a.a.) conservation across all the protein regions, including the well characterized bi-partite DBD, limiting our ability for inferring function from sequence (Fig. 1a and Extended Data Fig. 1a-b). Thus, we devised a mutagenesis strategy to systematically dissect OCT4 by sequentially deleting five a.a. stretches tiling across the entire OCT4 sequence with two a.a. overlaps, resulting in a total of 119X OCT4 derivatives (Extended Data Fig. 1c). These mutants were each inserted into a bacterial expression vector, generating 118 out of the 119 deletion mutants (del) as recombinant proteins with similar homogeneity and concentration to OCT4 wild-type (OCT4-WT) (Extended Data Fig. 1d).

To map the OCT4 segments essential for nucleosome binding, we compared the affinity of the recombinant OCT4-WT and the 118 mutants for a 162 base-pair (bp) DNA sequence derived from the *LIN28B* locus targeted by OSKM during reprogramming, either as naked DNA or as a reconstituted mono-nucleosomes using electrophoretic mobility shift assays (EMSAs) (Fig. 1b and Extended Data Fig. 2a-d)¹². Most deletions within the POU_S and POU_{HD} regions of OCT4-DBD (del-48-70 and del-79-95, respectively), with few exceptions

(see below), abolished OCT4 binding to both naked DNA and nucleosomes, indicating that the OCT4-DBD is essential for both naked DNA and nucleosome interactions (Fig. 1b and Extended Data Fig. 2b-c). Indeed, a recombinant OCT4-DBD can bind both DNA and nucleosomes (Fig. 1c and Extended Data Fig. 1e), albeit with reduced affinity, as indicated by significantly larger apparent dissociation constant (K_d) (Fig. 1d and Extended Data Fig. 1f). Furthermore, various deletions across the N-terminal (NT) and C-terminal (CT) regions of OCT4 reduced DNA and nucleosome binding (Fig. 1b and Extended Data Fig. 2a, b, d). Despite the reduced affinity, OCT4-DBD interaction with DNA and nucleosomes remains specific as only specific competitor (non-labeled DNA probes containing an OCT4-motif) can diminish this affinity (Fig. 1e). To rule out any deletion-specific effects, we also generated a sample of OCT4 mutants with equivalent alanine-stretch (ala.) substitutions scattered across different parts of OCT4 (Extended Data Fig. 1a and 3a). All OCT4-ala mutants displayed similar DNA and nucleosome binding activities as their deletion mutant counterparts (Extended Data Fig. 3b-c). In conclusion, specific nucleosome binding is encoded within the OCT4-DBD, which is strengthened by the NT and CT regions.

Uncoupling and enhancing OCT4 nucleosome binding

Remarkably, del-79 specifically abolished OCT4 binding to nucleosomes without significantly affecting affinity for naked DNA (dagger symbol in Fig. 1b and Extended Data Fig. 2c). By measuring the apparent K_d , we confirmed that OCT4-del-79 retained strong affinity for naked DNA, while binding to nucleosomes was almost completely abolished (Fig. 1c-d and Extended Data Fig. 1f). Furthermore, excess amounts of non-specific DNA competitor failed to compromise the interaction of OCT4-del-79 with *LIN28B*-DNA, demonstrating that this deletion did not affect the DNA-sequence specificity of OCT4 (Fig. 1e). Thus, despite being encoded within OCT4-DBD, nucleosome binding can be uncoupled from DNA binding through a deletion within OCT4-POU_{HD}.

Interestingly, OCT4-del-73 and del-75 within the linker region increased the affinity of OCT4 for both naked DNA and nucleosomes (asterisks in Fig. 1b and Extended Data Fig. 2c). For example, OCT4-del-73 mutant displayed more than two-fold lower K_d values for binding nucleosomes compared to OCT4-WT, indicating higher apparent affinity (Fig. 1c-d and Extended Data Fig. 1f). Furthermore, this OCT4 mutant showed similar DNA and nucleosome binding specificity to OCT4-WT (Fig. 1e). It is noteworthy that del-73 and del-75 contain E219 and E224 residues (homologous to E212 and E217 in mouse Oct4), which when mutated to alanine result in more affinity of mouse Oct4-DBD for DNA²⁹. Thus, nucleosome binding can be enhanced as well as uncoupled from DNA binding through distinct mutations.

Structural insights into OCT4-nucleosome binding

Although the three-dimensional (3D) structure of OCT4 co-bound with SOX2 to an engineered nucleosome has recently been solved using Cryo-EM, the linker and OCT4-POU_{HD}, which contain del-73 and del-79, respectively, were not observed²¹. This may be due to co-binding with SOX2 or the OCT4-site position in the engineered DNA sequence. We therefore used cryo-EM to solve the 3D structure of the *LIN28B*-nucleosome at an overall resolution of 3.5 Å (Fig. 2a-c and Extended Data Fig. 4a). DNA projecting

from the nucleosome core was also clearly visible in 2D averages (arrows in Fig. 2b). However, we observed a weak electron density at the 5' end compared to the 3' end (nucleosome entry/exit), indicating a higher degree of DNA flexibility (Fig. 2c). Importantly, we were able to resolve the DNA orientation within the *LIN28B*-nucleosome, revealing the position of all three possible OCT4 sites (Fig. 2d)^{12,30}. In site 1 and 3, the POU_S motif was solvent-exposed unlike the inaccessible POU_{HD} half-site, whereas only part of POU_{HD} half-site was exposed in site 2 (Fig. 2d). This observation indicates that OCT4 can target up to two partial motifs exposed on nucleosomes, while targeting at least three full motifs on naked DNA, explaining the multi-banding pattern observed with OCT4-DNA in EMSA compared to OCT4-nucleosome (Fig. 1c). We therefore modelled the binding of OCT4-DBD to these sites on the *LIN28B*-nucleosome. Modelling revealed that due to steric clashes, the OCT4 canonical DNA-binding configuration, where POU_{HD} and POU_S are on opposite sides of the DNA, is incompatible with nucleosome binding of OCT4 to any of these sites, including the most exposed site 1 at the entry/exit (Extended Data Fig. 4b-c). Canonical OCT4 DNA-binding could be accommodated at site 1 only under special circumstances where the flexible 5' end peels away from the nucleosome surface. However, OCT4 can also adopt a noncanonical configuration when bound to the more palindromic Oct factor recognition element (MORE) in which POU_{HD} and POU_S bind on the same side of DNA³¹. In this MORE configuration, OCT4 becomes compatible with binding the *LIN28B*-nucleosome on site 1 and 3 but not site 2 (Extended Data Fig. 4d-g), consistent with molecular dynamics simulations³². Interestingly, OCT4-POU_S binds specifically to the *LIN28B*-nucleosome whereas POU_{HD} is associated non-specifically in the MORE configuration, suggesting that POU_{HD} is contributing to the affinity but not specificity of OCT4-nucleosome interaction. This model suggests that residues within the del-79 region, and not so much the del-73 region, are in very close proximity to the DNA backbone and contacting the *LIN28B*-nucleosome (see red and brown regions in Fig. 2e). Thus, the OCT4 del-73 and del-79 regions might well contribute differently to nucleosome binding. Furthermore, OCT4 binding to a fully-wrapped *LIN28B*-nucleosome would involve a markedly different conformation and motif readout compared to the canonical OCT4 complex with naked DNA.

Next, we sought to evaluate the contribution of OCT4 interaction with histones to the overall nucleosome binding, which may be essential for binding nucleosomes but not naked DNA. Most proteins reported thus far interact with nucleosomes through the N-terminal tails of histones or the negatively charged binding interface on the surface of nucleosomes known as the acidic patch³³. In addition, crosslinking mass spectrometry (XL-MS) analysis has revealed that OCT4 may directly interact with the N-terminal tails of H2A, H3, and H4 when bound to *LIN28B* nucleosomes³⁰. We therefore generated five *LIN28B* nucleosome variants, each containing a tailless histone (H2A-tl, H2B-tl, H3-tl or H4-tl) or the acidic-patch (H2A-H2B) histone mutant (Fig. 2f-g). Interestingly, none of these histone mutants blocked the interaction of OCT4 with the *LIN28B*-nucleosome. Thus, the histone tails are not required for OCT4-nucleosome interaction but may be involved in nucleosome opening. Collectively, nucleosome-binding is uncoupled from DNA binding not through interrupting the interaction with histones but possibly by restricting the ability of OCT4 to adapt to nucleosomes.

Restricting OCT4 adaptation uncouples nucleosome binding

The five a.a. within del-79 (a.a. 236-240) are located within the N-terminal tail of POU_{HD}, each displaying various levels of conservation among POU TFs (Fig. 3a). To reveal the effects of a.a. 236-240 sequence on nucleosome-binding, we mutated these residues to alanine either individually (S236A, I237A, E238A, N239A, and R240A) or altogether (ala-79) as well as replacing the entire POU_{HD} of OCT4 with BRN2-POU_{HD} (BRN2HD) or BRN3-POU_{HD} (BRN3HD), which are members of the POU family but not involved in pluripotency (Fig. 3a and Extended Data Fig. 3d-e). Interestingly, apart from I237A and ala-79 that diminished OCT4 binding to nucleosomes and, to a lesser extent, to naked DNA, the other OCT4 mutants retained high affinity for both DNA and nucleosomes (Fig. 3b). These findings indicate that the del-79 region (a.a. 236-240) mediates interaction with nucleosomes across the POU family using various a.a. sequences, possibly by enabling OCT4 to adopt a conformation compatible with binding nucleosomal DNA rather than through specific interactions.

Thus, we investigated whether the length of del-79 region is the underlying cause of the observed loss of nucleosome binding. To that end, we incrementally deleted a.a. 236-239, generating OCT4-₁ to ₄ mutants that are missing 1 to 4 a.a. from the del-79 region, respectively (Fig. 3a Extended Data Fig. 3e). Unlike the mutations to alanine, the shortening of del-79 region even by one a.a. abolished nucleosome binding (Fig. 3c). However, binding to naked DNA was differentially affected by ₁ to ₄ mutations. Interestingly, shortening the del-79 region by 1 and 2 a.a. is more detrimental to binding DNA than 3, 4 or all 5 a.a. (Fig. 3c). Thus, shortening del-79 may affect the orientation of POU_{HD} relative to POU_S in a length-dependent manner, affecting the canonical OCT4 configuration to bind naked DNA as suggested by homology modelling (Fig. 3d). Such periodic effects were also observed with other mutants within POU_{HD}, including del-82-83 and del-86-87 displaying some DNA binding and del-80-81 and del-84-85 with no apparent DNA binding (double dagger symbol in Fig. 1b and Extended Data Fig. 2c). Therefore, our findings indicate that the distance between POU_S and POU_{HD}, not the a.a. sequence *per se*, mediates the ability of OCT4 to adapt to both DNA and nucleosome but in distinct ways.

Dynamic nucleosome binding shortens OCT4-SOX2 cooperativity

The relative affinity of TFs for DNA (thermodynamics) and the residence time of TFs on DNA (kinetics) both contribute to sequence-specific binding, but these properties can be modified independently from each other³⁴. We have identified more OCT4-del-73 in complex with DNA and nucleosome at equilibrium than OCT4-WT (Fig. 1c-d and Extended Data Fig. 1f). However, this apparent enhanced binding can be due to higher affinity or to faster on-off binding kinetics (shorter residence time). The caging effects of the gel matrix in EMSA can trap highly-dynamic bound and unbound DNA/nucleosome particles in the same retarded band, appearing as enhanced OCT4 binding. We therefore examined the average half-life ($t_{1/2}$) of OCT4-DNA/nucleosome complexes by challenging OCT4 binding with excessive amounts of specific competitor at increasing time intervals after equilibrium. We found that $t_{1/2}$ of OCT4-DNA complex is double that of OCT4-nucleosome for both WT and del-73 mutant, suggesting longer residence time of OCT4 on DNA compared to nucleosomes (Fig. 4a-c). Strikingly, the OCT4-del-73 mutant exhibits significantly faster

DNA as well as nucleosome binding kinetics ($t_{1/2}$ is one order of magnitude smaller) compared to wild-type despite the apparent enhanced binding at equilibrium (Fig. 4c). In summary, the binding of OCT4-del-73 to DNA and nucleosomes is enhanced by the kinetics and not the affinity of the interaction.

The cooperative binding of OCT4 and SOX2 to DNA has long been known to enhance stability and specificity²⁶. As expected, the OCT4-del-73 mutant displays enhanced cooperative binding with SOX2 to the *Fgf4* enhancer element, which contains the canonical OCT-SOX motif, compared to OCT4-WT (Fig. 4D). We thus sought to examine whether the dynamic binding behaviour of OCT4-del-73 remains when OCT4 is co-bound with SOX2. Again, the apparently enhanced OCT4-del-73-SOX2 cooperative binding is associated with significantly smaller $t_{1/2}$ compared to OCT4-SOX2, indicating faster binding dynamics (Fig. 4e-f). In summary, OCT4-del-73 displays accelerated binding kinetics to DNA and nucleosomes, resulting in short-lived OCT4-SOX2 co-operativity.

Nucleosome binding enables OCT4 to target closed chromatin

To further investigate the effects of nucleosome binding in cells, we carried out chromatin immunoprecipitation followed by sequencing (ChIP-seq) on OCT4-WT, del-73, and del-79 mutants after ectopic induction in mouse embryonic fibroblasts (MEFs) for 48h using a lentiviral dox-inducible system, showing comparable protein levels and chromatin fragmentation (Extended Data Fig. 5a-b). To measure the pre-existing chromatin accessibility (prior to OCT4 overexpression), we used ATAC-seq (assay for transposase-accessible chromatin using sequencing) in MEFs³⁵. As expected, OCT4-WT mainly targeted inaccessible chromatin regions (~ 64% closed), acting as a pioneer factor (Fig. 5a). In a striking contrast to OCT4-WT, OCT4-del-79 mutant was predominantly enriched at open chromatin sites (~ 71% open), recapitulating the ability to bind naked DNA but not nucleosomes *in vitro* (Fig. 5a). OCT4-del-73 however was enriched at more sites by comparison to OCT4-WT, especially within closed chromatin (Fig. 5a). The increased number of closed-sites occupied by OCT4-del-73 may relate to sampling more sites with shorter residence time that are fixed by prolonged (10 min) formaldehyde crosslinking. This is exemplified at the *Nanog* locus, which is essential for OCT4 function in reprogramming and pluripotency maintenance (Fig. 5b). Interestingly, OCT4-del-79 mutant showed the least overlap with OCT4-WT, diverging mostly within accessible chromatin, demonstrating that del-79 is more deficient in targeting closed chromatin (Fig. 5d). By contrast, the sites occupied by OCT4-del-73, particularly within closed chromatin, are mainly an expanded list of OCT4-WT sites (Fig. 5c). However, none of these mutations changed the distribution of OCT4 occupancy relative to the transcription start sites (TSS), which remained largely distant to TSS (Extended Data Fig. 5c). Therefore, uncoupling DNA binding from nucleosome binding drives OCT4 to preferentially target open chromatin, whereas shorter residence time on nucleosomes expands the ability of OCT4 to scan the genome within closed chromatin.

Interestingly, *de novo* motif analysis revealed that the closed sites targeted by OCT4 WT and del-73 are centrally enriched for a partial OCT4 motif, contrasting the open sites, which are almost depleted from any motif, suggesting that nucleosomes may enhance

OCT4-DNA binding specificity (Fig. 5d and Extended Data Fig. 5d-e). Moreover, no additional motifs relating to other TFs were identified in OCT4 WT and del-73 sites, consistent with individual nucleosome binding (Extended Data Fig. 5d-e). The open sites targeted by OCT4-del-79 however are enriched for other TF motifs such as JUNB:FOS and TEADs instead of OCT4 (Fig. 5d and Extended Data Fig. 5f). Therefore, OCT4-del-79 mutant relies on the specificity of other TFs to target open chromatin despite its ability to bind specifically to DNA *in vitro*. It is noteworthy that OCT4-del-79 closed sites, although representing the minority of sites, are also enriched for a homeodomain motif in addition to the JUNB:FOS motif, suggesting that OCT4-del-79 may be using its POU_{HD} or co-binding with other homeodomain TFs (Fig. 5d and Extended Data Fig. 5f). These findings suggest that OCT4 interaction with nucleosomes provide not only accessibility but also specificity to sites within closed chromatin.

OCT4 pioneer activity is required for reprogramming

To measure the functional contribution of the pioneer activity of OCT4 in reprogramming, we examined the ability of OCT4 WT, del-73 and del-79, each in combination with SKM, to reprogram MEFs to iPSCs, activating the pluripotency marker *Nanog* (Nanog positive colonies) (Fig. 6a). Both OCT4-del-73 and del-79 failed to induce pluripotency from MEFs, demonstrating the importance of stable nucleosome binding in reprogramming (Fig. 6b-c). Furthermore, the OCT4 alanine point-mutants that retained nucleosome binding retained the ability to induce pluripotency, unlike the I237A and ala-79 mutants and all deletion mutants with a shortened del-79 region and impaired nucleosome binding (Extended Data Fig. 6a-b). However, BRN2HD- and BRN3HD-OCT4 hybrids showed lower reprogramming activity despite their ability to bind nucleosomes (Extended Data Fig. 6a-b). To understand at what stage of reprogramming the BRN2HD and BRN3HD hybrids were most deficient, we expressed these factors with SKM for increasing periods of time (Extended Data Fig. 6c). Interestingly, both OCT4-BRN2HD and BRN3HD mutants display similar reprogramming efficiency to OCT4-WT during the initial 7 days, before they become most deficient at the end of reprogramming (Extended Data Fig. 6d-e). Therefore, a generic pioneer function within the POU_{HD} (BRN2HD and BRN3HD) is sufficient to initiate but not finalize reprogramming. Collectively, these data demonstrate that the affinity and the binding kinetics of OCT4 to nucleosomes are equally important for reprogramming.

Co-binding with SOX2 can rescue OCT4 pioneer deficiency

Next, we carried out CHIP-seq for OCT4 WT, del-73 and del-79 mutants when overexpressed with SKM in MEFs for 48h (OSKM-48h), an early stage of reprogramming (Fig. 6a). Surprisingly, all OCT4 del-variants, including the nucleosome-binding deficient OCT4-del-79, showed extensive closed chromatin occupancy (~ 80%) in the presence of SKM, unlike what was observed when expressed alone (compare Fig. 6d to Fig. 5a). Intriguingly, the OCT4-del-79 closed sites were solely enriched for SOX motifs rather than the OCT4 motif (Fig. 6e-f and Extended Data Fig. 7a), indicating that SOX2 may be the predominant driver for OCT4-del-79 to access closed chromatin. In contrast, the closed sites bound by OCT4-WT were mainly enriched for partial OCT4-motifs, in addition to the SOX2 motif as reported previously (Fig. 6e-f and Extended Data Fig. 7a)¹². Notably, the closed sites targeted by OCT4-del-73 were predominantly enriched for an OCT4 motif

but not SOX2, reinforcing the concept that dynamic nucleosome binding impairs the co-binding of OCT4 with SOX2 in early reprogramming (Fig. 6e-f and Extended Data Fig. 7a). Altogether, co-binding with SOX2 may rescue the pioneer deficiency of OCT4-del-79 mutant.

To validate the co-binding of OCT4 and SOX2, we carried out ChIP-seq for each of S, K, and M during early reprogramming and then isolated OCT4-unique sites (500bp or more away from any other S,K,M site) and OCT4-SOX2 co-bound sites (within 100bp range) (Fig. 6g). As expected, OCT4-unique sites are enriched for OCT4-WT and del-73 mutant, but not OCT4-del-79 (Fig. 6g). These OCT4-unique sites are also enriched for OCT4-WT and del-73 mutant, but not OCT4-del79, when expressed without SKM (Extended Data Fig. 7b). Importantly, OCT4-del-79 was enriched at sites co-bound by OCT4-SOX2 as exemplified in the *Nanog* locus (Fig. 6g-h). Furthermore, OCT4-del-79 is more enriched at SOX2-unique sites than OCT4-WT and del-73, suggesting that SOX2 can drive OCT4-del-79 even to its own sites (Extended Data Fig. 7c). Altogether, SOX2 can rescue the pioneer deficiency of OCT4, albeit at the expense of DNA-specificity, and therefore this is not sufficient to rescue reprogramming.

OCT4 pioneer activity is essential to maintain pluripotency

Once pluripotency is established in ESCs and iPSCs, most OCT4 sites are open, driving the expression of pluripotency genes. To investigate whether OCT4 pioneer activity is still required during pluripotency maintenance, we made use of the ZHBTc4.1 ES line carrying two null *Pou5f1* (gene that codes for Oct4) alleles and sustained by a *Pou5f1* transgene under the control of the tetracycline response element (TRE) using the Tet-Off system (Fig. 7a)³⁶. First, we examined whether OCT4-WT, del-73, del-79 as well as the other del-79 derivatives can maintain ZHBTc4.1 ESCs in an undifferentiated state (Alkaline Phosphatase (AP) positive) after knocking down the *Pou5f1* transgene by administration of Dox (Fig. 7a-b). Unlike OCT4-WT, which fully rescued pluripotency in ZHBTc4.1 ESCs after Dox treatment³⁷, OCT4-del-73 failed to maintain these cells in an undifferentiated state (Fig. 7b-c). Thus, the highly dynamic binding of OCT4-del-73 to DNA and nucleosomes may adversely impact OCT4-SOX2 co-operativity in ESCs, which is known to be essential for pluripotency⁴. Additionally, the OCT4-del-79 mutant and all del-79 derivatives that are deficient in binding nucleosomes were also deficient in maintaining pluripotency when ZHBTc4.1 ESCs were treated with Dox (Fig. 7b-c and Extended Data Fig. 8a). In contrast, alanine substitutions within the del-79 region that preserved OCT4 binding to nucleosomes also retained the ability of OCT4 to maintain pluripotency in ZHBTc4.1 ESCs (Fig. 7b-c and Extended Data Fig. 8a). Together, these findings indicate that continuous and stable targeting of nucleosomes are required to act against forces destabilising pluripotency.

By contrast, the OCT4-BRN2HD and BRN3HD hybrids displayed diminished capacity to maintain pluripotency, which is important at the final stages of reprogramming (Extended Data Fig. 8b-c). This could explain why these mutants fail to finalize reprogramming. Thus, while nucleosome binding can be interchanged amongst various POU-homeodomains, full reprogramming and pluripotency maintenance require the specific OCT4-POU_{HD} beyond the del-79 region.

Open enhancers involves sustained OCT4-nucleosome binding

To shed light on the mechanism by which OCT4 pioneer activity maintains pluripotency, we investigated the effects of OCT4-del-73 and del-79 mutants on chromatin accessibility in ESCs (Fig. 7a). We therefore carried out ATAC-seq in ZHBTc4.1 ESCs transfected with one of the OCT4 variants prior to Dox treatment (0h), representing the initial pluripotency state, and 24h after Dox treatment at which time the endogenous Oct4 has been fully replaced by the exogenous counterpart (Extended Data Fig. 8d-e). In addition, we carried out ChIP-seq for endogenous Oct4 in ESCs and identified sites associated with pluripotency that are predominantly within open chromatin (Fig. 7d). Remarkably, in the absence of endogenous Oct4 (24h), the accessibility of the majority of open sites (n=52,774) was significantly diminished (closed down) in the presence of exogenous OCT4-del-73 and del-79 mutants but not OCT4-WT (Fig. 7d-e). Nonetheless, there was no change in chromatin accessibility (remained open) for the rest of the Oct4 open sites (n=35,928) at 24h in all exogenous OCT4 conditions (Fig. 7d-e). Moreover, the Oct4 closed sites (n=28,378) remained inaccessible (remained closed) with all exogenous OCT4 variants (Fig. 7d-e). We then measured nucleosome enrichment at these three groups of Oct4 sites using Micrococcal nuclease sequencing (MNase-seq) data previously generated in mESCs¹³. Remarkably, the Oct4 sites that closed down were significantly enriched for nucleosomes, indicating that Oct4 is bound to nucleosome even within accessible chromatin (Figure 7d and 7f). This is in contrast to sites that remained open, which are depleted of nucleosomes (Figure 7d and 7f). Oct4 sites that remain closed are also enriched for nucleosomes although less central and spread across the flanking regions, reflecting the inaccessible chromatin state (Figure 7d and 7f). Hence, stable OCT4-nucleosome binding is required to maintain open chromatin in ESCs.

Interestingly, the OCT4 sites that closed down are predominantly distal from TSS, reminiscent of enhancers, like the enhancer of the Undifferentiated Embryonic Cell Transcription Factor 1 (*Utf1*) gene (Fig. 7g-h). Conversely, the OCT4 open sites that remained open are mainly proximal to TSS, acting as promoters such as that of Laminin Subunit Alpha 5 (*Lama5*) (Fig. 7g-h). The closed sites however are equally distributed around TSS, representing both enhancers and promoters (Fig. 7g-h). Thus, a stable interaction of OCT4 with nucleosome is specifically required to maintain the accessibility of open enhancers in ESCs, suggesting that maintaining the accessibility of promoters may rely on other factors.

Additionally, the gene ontology (GO) annotations of genes associated with Oct4 closed down sites were limited to the regulation of stem cell population maintenance, and cellular response to leukaemia inhibitory factor (Extended Data Fig. 8f)³⁸. However, the GO terms of genes targeted by Oct4 sites that remained open are more diverse, ranging from regulation of chromatin and gene expression to mitochondria membrane and cellular adhesion (Extended Data Fig. 8f). The GO terms associated with OCT4 sites that remained closed with all OCT4 variants belong to heterochromatin organization, mesodermal cell fate commitment, and mitotic sister chromatid cohesion (Extended Data Fig. 8f). Therefore, our data demonstrate that the accessibility of pluripotency-maintenance enhancers is particularly dependent on the stable interaction of OCT4 with nucleosomes.

Discussion

The interaction of TFs with the genome, which is interlaced with nucleosomes, plays a central role in determining cell fate. Yet, nucleosomes have long been considered as barriers to TF-DNA binding^{39,40}. Furthermore, pioneer TFs can bind nucleosomes by recognizing partial DNA motifs exposed on the nucleosome surface, correlating with access to closed chromatin^{5,12}. However, the functional contribution of TF-nucleosome binding to cell identity remains poorly defined. This gap in our knowledge led our effort to systematically map the interaction OCT4 with nucleosomes. In doing so, we also revealed that OCT4 nucleosome-binding can be: (i) uncoupled from DNA-binding and (ii) enhanced by accelerated DNA-binding. Exploiting this ability to manipulate OCT4-pioneer activity has shown that nucleosomes play a supportive role, rather than an obstructive role, in facilitating not only chromatin accessibility but also in obtaining specific TF-DNA interaction to engage the genome.

It is conceivable that OCT4 nucleosome-binding can mediate access to closed chromatin, targeting silent genes during reprogramming¹. However, it was unclear how nucleosome-binding can maintain open chromatin, which is generally regarded as depleted of nucleosomes. Although, this view has been challenged, as various studies have shown that accessible sites bound by Oct4 and other pioneer TFs are still enriched for nucleosomes^{17,41–43}. Interestingly, nucleosomes within open chromatin are more sensitive to nuclease digestion and may appear as depleted if over-digested^{13,17,41–43}. Our data demonstrate that a stable OCT4-nucleosome interaction is specifically required to maintain open enhancers, suggesting that opening chromatin may lead to nucleosome reconfiguration that involve chromatin modifiers without totally evicting nucleosomes^{2,17,41,43–45}. Intriguingly, the pioneer function of OCT4 is redundant in maintaining the accessibility of promoters in ESCs, which appear to be depleted of nucleosomes. Therefore, studying the distinct mechanisms that establish and maintain open enhancers versus promoters may reveal more details about the role of nucleosomes in genome accessibility.

Methods

Animals

Primary MEFs were generated from 129-mouse embryos at E12.5. On average 6-8 embryos were collected from one pregnant female, resulting in 4×10^6 cells per embryo. The sex of embryos were determined to generate MEFs. All animal experiments were approved by the University of Edinburgh Animal Welfare and Ethical Review Body, performed at the University of Edinburgh, and carried out according to regulations specified by the Home Office and Project License.

DNA and plasmid construction

The OCT4 deletion library, alanine point mutants, and the homeodomain hybrids were synthesized by Twist Bioscience (San Francisco, CA) and supplied as linear DNA fragments (Supplementary Table 1 and 2), which were individually cloned into pET28b (Novagen, Merck Millipore, Burlington, MA) for expression in *E. coli*. Cloning was achieved by

PCR amplification of each linear fragment of DNA using forward and reverse primers that included an EcoRI and NotI site, respectively (Supplementary Table 3). The amplified PCR products were digested EcoRI and NotI and then ligated into a linearized pET28B vector. The alanine stretch OCT4 mutants were generated by site-directed mutagenesis of pET28b-OCT4 plasmid using the QuikChange II kit (Stratagene) following the manufacturer's instructions. The incremental deletions OCT4- 1- 4 mutants and the OCT4-ala-79 mutant were generated as synthetic DNA fragments and inserted into pET28A between EcoRI and NotI sites by Twist Bioscience (San Francisco, CA) (Supplementary Table 2).

For reprogramming studies, the OCT4 mutants were PCR amplified using a set of primers containing EcoRI and XbaI (Supplementary Table 2) and the correspondent pET28-OCT4 plasmids as templates. The amplified products were digested with EcoRI and XbaI and cloned into linearized FUW-tet-O-XbaI plasmid, which was modified from FUW-tet-O-hOct4 plasmid (Addgene plasmid; #20726)⁴⁶, by engineering an XbaI site to replace the EcoRI site (nucleotides 4139 to 4144) using a QuikChange II site-directed mutagenesis kit (Stratagene, La Jolla, CA). The lentivirus plasmids FUW-tet-O-hOct4 (Addgene plasmid #20726), FUW-tet-O-hSox2 (Addgene; plasmid #20724), FUW-tet-O-hKlf4 (Addgene; plasmid #20725) and FUW-tet-O-hMyc (Addgene; plasmid #20723) encoding for OSKM, respectively, were generated in the Rudolf Jaenisch laboratory⁴⁶. The pWPT-rtTA2M2 vector was generated in the Kenneth Zaret laboratory⁴⁷. For pluripotency rescue experiments, the OCT4 mutants were PCR-amplified from the pET28-OCT4 constructs and sub-cloned into pCAG-IRES-Puro³⁶ using the In-Fusion HD Cloning Plus kit according to the manufacturer's instructions (Takara Clonotech, Tokyo, Japan).

The N-terminal tail-less human histone mutants were generated by PCR, removing the coding region of AA 2-14 from H2A, AA 2-32 from H2B and cloned into the pET15B plasmid. The N-terminal tail-less human histone H3 and H4 and the acidic patch mutants H2A and H2B were a gift from Dr Marcus Wilson lab at the University of Edinburgh. All the wild-type human histones were inserted into pET15B vector as described previously¹².

Protein expression and purification

Recombinant proteins were expressed in *E. coli* by transforming Rosetta™ 2 (pLysS) host strains (Novagen, Merck Millipore) with each of the recombinant pET28b constructs encoding wild-type hOCT4 or a mutant version of the protein fused to an N-terminal 6 x histidine tag. Cells were cultured at 37°C in 5mL of LB medium supplemented with chloramphenicol and kanamycin to an optical density at 600nm of ~0.4. Heterologous gene expression was then induced by addition of IPTG to a final concentration of 0.5mM and growth was continued at 37°C for a further 5h. The harvested cell pellets from a 5mL culture were resuspended in PBS containing cComplete™ EDTA-free Protease Inhibitor Cocktail (Roche, Basel). The cells were then disrupted by sonication using a Bioruptor™ sonicator (Diagenode, Liège) on high power setting (10X 30s bursts with 30s intervals). Sonication was performed in a cold room (4°C) using chilled water in the sonication bath. The cell lysate was clarified by centrifugation at 5000g for 20min at 4°C and the pellet (insoluble fraction) collected. Subsequent purification steps were carried out at room temperature. The insoluble fraction was resuspended in 600µl of denaturing buffer (DB) comprising: 8M urea,

0.1M NaH₂PO₄ (pH8.0), 0.3M NaCl, 10mM imidazole and then applied to a HisSpinTrap column (GE Healthcare, Chicago) pre-equilibrated in the same buffer. Sample application and washing was carried out according to the manufacturer's instructions. After washing the column with DB, bound proteins were eluted in 200µl elution buffer (EB) (same as DB but containing 500mM imidazole).

The purified denatured proteins were refolded by buffer exchange to remove EB and replace with refolding (RF) buffer (50mM MES (pH5.5), 240mM NaCl, 10mM KCl, 2mM MgCl₂, 2mM CaCl₂, 0.8M urea, 30% (v/v) glycerol, 0.1% NP40 substitute, 0.05% (v/v) Triton X100, 2mM EDTA, 5mM DTT). Buffer exchange was performed using a PD-SpinTrap G25 column (GE Healthcare) equilibrated in RF buffer according to the manufacturer's instructions. A 180µl aliquot of purified protein from the HisTrap column was applied to the PD-SpinTrap G25 column. The purified protein was aliquoted and snap frozen using liquid nitrogen before storage at -80°C. Protein purity was assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using the Bolt system with 4-12% gradient Bis-Tris gels (Thermo Fisher Scientific, Waltham, MA), followed by staining using GelCode™ to visualise protein bands (Thermo Fisher Scientific). Protein quantification was estimated by densitometry measurements of protein bands in digital images of SDS-PAGE gels stained with GelCode using Multi Gauge image software ver2.0 (FujiFilm, Minato, Tokyo).

Nucleosome reconstitution

Cy5-labelled *LIN28B* DNA was amplified from pUC19 plasmid containing a 162bp DNA sequence from the *LIN28B* locus¹² using the following Cy5-end-labelled primers (Integrated DNA Technologies, Coralville, IA):

5'-AGTGGTATTAACATATCCTCAGTGGTG-3'

5'-TGTCTTTATTACACAAGCTTGCACAA-3'

The amplified PCR product was gel extracted using a Qiagen gel extraction kit (Qiagen, Hilden, Germany) and then further purified by anion exchange chromatography using a Mono Q 5/50 GL column (GE Healthcare).

Small scale nucleosome assembly for EMSA was performed using the Cy5-labelled *LIN28B* DNA and recombinant human histones refolded to H2A/H2B dimers and H3/H4 tetramers as described previously¹². Briefly, a 500µl mixture containing 6µg H2A/H2B, 6µg H3/H4 and 10µg Cy5-labelled *LIN28B* DNA in 4M urea, 2M NaCl and 1mg/mL BSA was prepared. The mixture was dialyzed against 1L denaturing buffer containing 5M urea, 10mM Tris HCl pH8.0, 1mM EDTA, 2M NaCl, and 10mM 2-mercaptoethanol at 4°C overnight. The salt level in the sample was then gradually reduced by successive dialysis steps at 4°C against 1L of the same denaturing buffer containing 1.5M NaCl for 2hr, 1M NaCl for 2hr, 0.8M NaCl for 2hr and then 0.6M NaCl overnight. Next, the sample was dialyzed against 1L of non-denaturing buffer containing 10mM Tris HCl pH8.0, 1mM EDTA, 0.6M NaCl, 1mM 2-mercaptoethanol at 4°C for 6hr. The sample was then dialyzed against the non-denaturing buffer containing 0.1M NaCl at 4°C overnight. The nucleosome

preparation was centrifuged (15,000g at 4°C for 10min) and the pellet discarded. Finally, the nucleosomes were heat-shifted by incubation at 37°C for 6hr. An aliquot of the nucleosome preparation was analysed on native 5% polyacrylamide gel alongside free *LIN28B* DNA to assess its quality and quantity by Ethidium Bromide staining and Cy5-fluorescence. Large scale nucleosome reconstitution for Cryo-EM studies was carried out using histone octamers as previously described⁴⁸.

Electrophoretic Mobility Shift Assays

The binding to Cy5 end-labelled *LIN28B*-DNA/nucleosome or oligonucleotide duplexes were analysed in native 4% or 5% polyacrylamide gels (dimensions: 0.15 x18x18cm), respectively, which were prepared in 0.5X TBE (45mM Tris-borate, 1mM EDTA). Gels were stored overnight at 4°C in 100% humidity before pre-running at 90V (approximately 10V/cm) for 1hr. For affinity analysis, a 40µl mixture typically containing 1nM Cy5-labelled DNA or nucleosome and 0 to 8nM purified wild-type or mutant OCT4 was prepared in 1X binding buffer (10mM Tris HCl pH7.5, 1mM MgCl₂, 10µM ZnCl₂, 10mM KCl, 1mM DTT, 5% (v/v) glycerol, 0.5mg/mL BSA). The mixtures were incubated at 20°C ±1°C in the dark for 1hr using protein LoBind tubes (Eppendorf UK Ltd). A 30µl aliquot of each sample was then loaded onto acrylamide gels and electrophoresis was performed at 90V for 4hr. The gels were imaged by detecting Cy5-fluorescence using a FujiFilm Life Science FLA-5100 instrument (FujiFilm) or BioRAD ChemiDoc MP imaging system. The resulting images were visualized, and the bands quantified using Multi Gauge image software ver2.0 (FujiFilm).

For competition assays, the experiments were conducted in the presence or absence of a 40-fold excess of unlabelled competitor DNA (either specific DNA containing the cognate OCT4 binding site from the *FGF4* promoter or non-specific DNA lacking a binding site) over 2nM Cy5-labelled DNA (see Supplementary Table 3 for sequences).

Experiments were also carried out to assess the decay of a preformed complex of OCT4-WT or mutants with Cy5-labelled DNA or nucleosomes in the presence of a 40-fold excess amount of unlabelled specific competitor DNA. Here, a master mix of the complex was first generated and then an aliquot loaded onto the gel, which represented the zero timepoint. A concentrated solution of unlabelled specific competitor DNA was then added to the remaining master mix to give a 40-fold excess. The master mix was incubated at 20°C ±1°C and aliquots of this mixture were subsequently loaded onto the gel at specified timepoints. Equivalent decay experiments were conducted using preformed complexes of OCT4/SOX2 with Cy5-labelled FGF4 DNA. Such complexes were generated by adding a mixture of OCT4 and SOX2 to an equal volume of Cy5-labelled FGF4 DNA and then incubating the resulting solution at 20°C ±1°C in the dark for 1hr prior to addition of competitor DNA. The apparent dissociation constant and half-life values were calculated as explained in the Supplementary protocols.

Cryo-EM sample preparation and data collection

Cryo-EM grids were prepared using a 0.5mg/ml sample of *LIN28B*-nucleosome core particles (4µL) applied onto 200 mesh Quantifoil grids with 2µm holes and 2µm spacing

(Quantifoil Micro Tools GrmbH). The sample was blotted using a Vitrobot Mark IV (Thermo Fisher Scientific) operated at 4°C with 2.5s blot time, -1 blot force, 10s wait time and 100% humidity before plunge freezing in liquid ethane. The frozen grids were stored in liquid nitrogen until imaging.

Cryo-EM data collection was performed at Birkbeck, University of London using a Titan Krios electron microscope (Thermo Fisher Scientific) operating at 300keV and equipped with a BioQuantum K3 energy filter (20eV slit width) and K3 summit direct electron detector. Data were collected using the K3 in super resolution mode at a nominal magnification of 130kx (0.335 Å per super resolution pixel). Movies composed of 50 frames (0.03s/frame) were collected at a dose rate of 16.814 e-/pixel/sec corresponding to a total accumulated dose of 56.7 e-/Å². A dataset of 12892 movies was collected using defocus values ranging from -1.5 to -3.3µm with 6x exposures per hole (0.8µm illumination area) using the EPU software package (Thermo Fisher Scientific). Cryo-EM data collection statistics are summarized in Extended Data Fig. 4a. Cryo-EM data processing and model building are indicated in the Supplementary protocols.

Lentivirus Production and titration

To produce each of OCT4-WT or mutants, SOX2, KLF4, c-MYC and rtTA2M2 lentiviral supernatant, nearly one million HEK 293T cells were cultured for 24 hr in GMEM medium supplemented with 10% Foetal Calf Serum (FCS), 1 mM sodium pyruvate, 1 mM glutamine, and Non-Essential Amino Acids at 37°C and 5% CO₂. Then at ~40% confluency, the HEK 293T cells were transfected with a plasmid cocktail containing 2.5 µg expression plasmid, 1.7µg psPAX2 packaging vector, 0.8µg envelope vector mixed in 30µl Fugene6 (Roche) and 570µl OPTI-MEM medium (Invitrogen, Thermo Fisher Scientific). The cells were further cultured for 55hr at 37°C and 5% CO₂ before the virus-containing supernatant was harvested, cleared by centrifugation, and filtered through a 0.45µm syringe filter (Merck Millipore). Final volume of recovered viral supernatant was measured and polybrene was added to a final concentration of 4.5µg/ml. The viral supernatant was aliquoted and stored at -80°C.

To produce concentrated virus for ChIP-seq experiments, lentivirus supernatant was generated from 3million HEK 293T cells as described above, which was then transferred to thin-walled ultracentrifuge tubes and pelleted by ultracentrifugation at 25,000 rpm using a SW32-TI rotor in a Beckman Optima XPN ultracentrifuge (Beckman Coulter, Brea, CA). The pelleted virus was resuspended in 100µl plain GMEM medium and incubated at 4°C for 16hr. The viral titre calculated by flow cytometry was 5x10⁸ infectious units per ml.

Reprogramming MEFs to iPSCs

Early passage (passage-1) mouse embryonic fibroblasts (MEFs) were cultured for 48hr in MEF medium (GMEM medium supplemented with 10% FCS, 1 mM sodium pyruvate, 1 mM glutamine, and 5.7ml MEM 100X Non-Essential Amino Acids Solution, 50µM 2-mercaptoethanol (Gibco, Thermo Fisher Scientific) at 37°C and 5% CO₂. Twenty-four hours prior to viral transduction, 40,000 cells were seeded on gelatinized 6-well plates. The cells were transduced by exchanging the growth culture media with lentiviral media

containing an equal volume mixture of OCT4-WT or mutant along with SOX2, KLF4, c-MYC and rtTA2M2 viral supernatants. Viral media were replaced with fresh MEF medium at 8 hr post transduction. At 24 hr post transduction, the culture medium was replaced with reprogramming medium containing LIF and 1 $\mu\text{g/ml}$ doxycycline (induction day-0). The medium was renewed every two days until induction day-6, whereupon it was renewed daily. At induction day-10, doxycycline was removed from the reprogramming medium and emerging iPSC-like colonies were further cultured for 4 days in the absence of doxycycline.

For reprogramming efficiency quantification, iPSC colonies were fixed with 4% paraformaldehyde for 10min at room temperature on induction day-14. Fixed cells were washed twice with rinse buffer (20mM Tris-HCl pH 7.4, 150mM NaCl, 0.05% Tween-20), permeabilized with 0.1% Triton X-100 for 10min at room temperature and blocked with 4% Donkey serum (Sigma-Aldrich) in PBS for at least 60min at room temperature. Blocked cells were incubated overnight in blocking buffer containing 0.5 $\mu\text{g/ml}$ anti-mouse Nanog antibody (eBioscience). Antibody-stained cells were washed three times with PBS containing 0.05% Tween-20. After washing, cells were incubated with 1 $\mu\text{g/ml}$ Donkey Anti-Rat IgG H&L-Alexa-Fluor-488 (Invitrogen, Thermo Fisher Scientific) in PBS for 2hr at room temperature. Nuclei were stained with 3mg/ml DAPI for 10min at room temperature. Whole well containing iPSC colonies were imaged at a resolution of 4 $\mu\text{m/pixel}$ using a CELIGO image cytometer (Nexcelom, Lawrence, MA). Colonies with DAPI signal covering an area larger than 8 adjacent pixels were counted. Nanog-positive colonies were counted if at least 10% of the DAPI signal overlapped with the Nanog signal. To eliminate auto-fluorescence, images were also taken using the red channel. Colonies were counted in ImageJ software⁴⁹.

Pluripotency maintenance and rescue

ZHBTc4.1 ES cells were generated in the Austin Smith laboratory as described previously³⁷. ZHBTc4.1 ES cells were maintained in GMEM supplemented with 10% FCS, 1 mM glutamine/pyruvate, Non-essential Amino Acids, Penicillin/Streptomycin and 0.05 mM 2-mercaptoethanol on porcine gelatine-coated 6-well plates. Nearly one million cells were seeded and transfected with 2.5 μg plasmid DNA using Lipofectamine 3000 (Thermo Fisher Scientific) diluted in OptiMEM reduced serum medium (Thermo Fisher Scientific) following the protocol recommended by the manufacturer. For *Pou5f1* repression by doxycycline (24hr post-transfection), 100,000 ZHBTc4.1 ES cells were harvested and cultured for 4hr before treating with doxycycline (1 $\mu\text{g/ml}$) and puromycin (2 $\mu\text{g/ml}$). Cells not treated with doxycycline were used as -Dox control. The medium was then replaced every 48hr for a total of three changes. On day-7 of culture, colonies were fixed and stained for the presence Alkaline Phosphatase (AP) (Sigma-Aldrich) according to the manufacturer's instructions. AP positive colonies were manually counted using a Stereo Microscopes (Olympus, Tokyo, Japan). Colony staining and cell morphology were used as criteria for pluripotent ESCs. The Rescue Index was calculated as the ratio of the mean +Dox counts over the mean of -Dox counts. Error bars represent the standard deviation from three replicates.

Western Blot Analysis

Whole cell extracts were prepared from DOX-induced and uninduced MEFs using RIPA extraction buffer (25 mM Tris HCl pH 7.5, 150 mM NaCl, 1% Na-deoxycholate, 1% NP-40, 0.1% SDS) supplemented with Complete protease inhibitor cocktail (Roche) and Halt phosphatase inhibitor cocktail (Thermo Fisher Scientific). The protein concentrations of the lysates were quantified using Pierce™ BCA Protein Assay Kit according to the manufacturer's instructions (Thermo Fisher Scientific). Proteins resolved by SDS-polyacrylamide gel electrophoresis were electroblotted onto a PVDF membrane. The primary antibody incubations with anti-human OCT4 antibody (0.45mg/ml #ab19857; Abcam) and anti-human GAPDH (1:5000 dilution of GTX627408; GeneTex Inc, Irvine, CA) were performed at 4°C for 16hr. The secondary antibody incubations with goat anti-rabbit IgG-HRP (1:5000 dilution of #sc-2004; Santa-Cruz Biotechnology, Santa-Cruz, CA) and goat anti-mouse IgG-HRP (1:5000 of #sc-2005 Santa Cruz Biotechnology) were performed for 1hr at room temperature. Blots were visualized by using SuperSignal™ West Pico Chemiluminescent Substrate (Thermo Fisher Scientific) using Amersham Hyperfilm ECL (GE-Healthcare) developed in Mi5 Processor (Jet X-Ray, London) or BioRAD ChemiDoc MP imaging system.

Chromatin Immunoprecipitation (ChIP-seq)

Early passage MEFs (5×10^6 cells) were seeded onto 15-cm dishes (3 dishes in total) and cultured for 24 hr in MEF medium before being infected with either Tet-on-O (WT or mutant) and rtTA2M2 or Tet-on-O (WT or mutant), -S, -K, -M and rtTA2M2 lentiviruses at a multiplicity of infection (M.O.I.) of 5 for each virus. OCT4 or OSKM expression was induced by treating infected cells with 1µg/mL Dox for 48hr. Cells were then cross-linked with 1% formaldehyde for 10min at room temperature. Crosslinking was quenched by adding 120mM glycine and incubating for 10min at room temperature. Cells were harvested using a scraper and pelleted by centrifugation at 1350g for 5min at 4°C. The crosslinked pellet was washed 3 times with 10ml of ice-cold PBS before flash frozen in liquid nitrogen and stored at -80°C.

Before nuclear extraction, the cell pellet was thawed on ice for 2-3hr and resuspended in 10ml filtered, ice-cold lysis buffer (50mM HEPES-KOH pH7.5, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100 and 1 tablet of Complete Ultra protease inhibitor cocktail (Roche)). The suspension was gently mixed on a rotating platform at 4°C for 10min. The cells were disrupted using a 7ml glass-Dounce homogeniser (40 strokes) on ice. The nuclei were pelleted by centrifugation (1350g for 5min at 4°C) and washed with 10ml ice-cold wash buffer (10mM Tris-HCl pH 8, 200mM NaCl, 1mM EDTA, 0.5mM EGTA and 1 tablet of Complete ultra-protease inhibitor cocktail (Roche)) for 10min at 4°C. The nuclei were collected by centrifugation and resuspended in 4ml sonication buffer (10mM Tris-HCl pH 8, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% Na-deoxycholate, 0.5% N-lauroylsarcosine and 1 tablet of Complete ultra-protease inhibitor cocktail (Roche)). The resuspended nuclei were split into four aliquots in pre-chilled 1mL milli-tubes containing AFA Fibre and sonicated using a Covaris-M220 Focused-ultrasonicator™ (Covaris, Woburn, MA). Each milli-tube was sonicated for 10min and kept on ice for 30min per sonication cycle for a total of 7 cycles. Sonicated chromatin

was transferred to Protein-Lobind tubes (Eppendorf). 100µL of 10% Triton X-100 was then added to each 1ml sonicated chromatin to increase solubility. Chromatin samples were then centrifuged (20,000g at 4°C for 10min) and the supernatants pooled into fresh tubes. A 50µL aliquot of each pooled sample was analysed to check the fragment size distribution and quantify the DNA content of the resulting sonicated chromatin using a Nanodrop spectrophotometer. Another 50µL aliquot was retained to be used as input for ChIP analysis. The sonicated chromatin and the input sample were snap-frozen in liquid nitrogen and stored at -80°C.

For each ChIP, 30µl Dyna Protein-G magnetic beads (Thermo Fisher Scientific) were washed three times with 1ml blocking solution (0.5% w/v BSA in PBS-Tween-20). The beads were saturated with the appropriate antibody by adding 10µg of either OCT4 (Abcam), SOX2, KLF4 or cMYC (R&D) antibody and incubated on a rotating platform at 4°C for at least 6hr. Sonicated chromatin (40µg) was then mixed with the antibody-saturated beads and incubated on a rotating platform at 4°C overnight. The beads were transferred to a fresh tube, washed 5 times with 1mL wash buffer (50mM HEPES-KOH pH7.6, 500mM LiCl, 1mM EDTA, 1% NP-40, 0.7% Na-deoxycholate) and then washed once more with TE buffer (10mM Tris-HCl pH8, 1mM EDTA) containing 50mM NaCl. The bound-chromatin was eluted by incubating the beads in 210µl elution buffer (50mM Tris-HCl pH8, 10mM EDTA, 1% SDS) at 65°C for 30min. The beads were pelleted by centrifugation at 16000g for 1min and 200µl supernatant containing the soluble chromatin was transferred to fresh tubes. The crosslinking was reversed by incubating the eluted chromatin and input DNA (in elution buffer) at 65°C for 16hr with shaking. An equal volume of TE (200µl) was added to the eluted chromatin and input DNA to reduce the concentration of SDS. RNA in the samples were digested by adding 0.2mg/ml RNase-A for 2hr at 37°C. Proteins were then digested by adding 0.2mg/ml Proteinase-K for 2hr at 55°C. ChIP and Input-DNA was subsequently purified by phenol-chloroform extraction followed by ethanol precipitation and resuspended in EB buffer (100 mM Tris pH 8), before DNA concentrations were measured by Qubit Fluorometric Quantitation (Invitrogen, Thermo Fisher Scientific).

ChIPed DNA from three biological replicates were pooled for each OCT4 (WT or mutants), SOX2, KLF4, MYC and Input controls. The amount of DNA used in the library preparation ranged from ~ 20–80ng. Input-DNA from each sample was pooled in equal amounts for a single library preparation. The DNA libraries were prepared using NEBNext Ultra-II DNA Library Prep-Kit (New England Biolabs, Ipswich, MA) according to the manufacturer's instructions. Each library was uniquely barcoded using NEBNext® Multiplex Oligos for Illumina® (Dual-Index Primers Set-1) (New England Biolabs). Size selection of the DNA libraries was carried out using SpeedBeads™, which comprise magnetic carboxylate modified particles (Thermo Fisher Scientific). The quality of the DNA libraries was assessed using Agilent HS-DNA-Screen Tape (Agilent, Santa Clara, CA). The DNA libraries were then pooled by mixing an equal volume of each ChIP DNA library. The concentration of the DNA library pool was determined to be around 3nM by Qubit Fluorometric Quantitation. The sequencing of the OCT4-48h ChIP-seq libraries were carried out by Edinburgh Genomics in Illumina NovaSeq and OSKM-48h ChIP-seq sequencing was done by Beijing Genomics Institute using Illumina HiSeq-4000 system, both using a 50bp pair-end next generation sequencing (Illumina, San Diego, CA).

Chromatin accessibility (ATAC-seq)

Two sets (untreated and dox-treated) of one million ZHBTc4.1 ES cells were transfected with 2.5 µg circular pCAG plasmid encoding for OCT4-WT or mutant using Lipofectamine-3000 (Thermo Fisher Scientific) following the protocol recommended by the manufacturer. After 24hr, both sets were cultured in selection media containing puromycin (2µg/ml) for 2 days to eliminate non-transfected cells. Cells of the untreated set were harvested, counted, and immediately used for the ATAC-seq protocol (see below). Doxycycline (1µg/ml) was added to the culture medium of the treated set for 24hr to repress *Pou5f1* transgene. Cells of the dox-treated set were then harvested, counted, and immediately used for ATAC-seq protocol.

Around 50,000 of freshly harvested cells were resuspended in ice-cold RSB buffer (10mM Tris-HCl pH7.4, 10mM NaCl, 3mM MgCl₂). The suspension was centrifuged at 500xg for 3min using a refrigerated centrifuge. The cell pellet was washed once with 50µl ice-cold PBS and centrifuged at 500xg for 3min at 4°C. The cell pellet was resuspended in ice-cold Lysis Buffer (48.5µl RSB buffer supplemented with 0.5µl of 10% NP-40, 0.5µl of 10% Tween-20, 0.5µl of 1% Digitonin) and kept on ice for 3min. Then, 1ml of wash buffer (RSB buffer supplemented with 0.1% Tween-20) was added on the lysed cells. Nuclei were pelleted at 500xg for 10min at 4°C and the supernatant (cytoplasm fraction) was carefully discarded. The nuclei were resuspended in the transposition mix (25µl of 2x Tagment DNA Buffer (Illumina, #20034197), 16.5 µl of 1X PBS, 0.5 µl 10% of Tween-20, 0.5 µl of 1% Digitonin, 5µl nuclease-free H₂O). Next, 2.5µl of Tn5 Transposase (Illumina, #20034197) was added to the reaction mix by pipetting up and down 6 times. The reaction was then incubated at 37°C for 30min on a thermomixer at 1,000rpm. Finally, the tagmented DNA was isolated using Qiagen MinElute Reaction Cleanup-Kit (Qiagen, #28204) following the manufacturer's instructions.

ATAC-seq libraries were generated by PCR amplification using custom-made barcodes⁵⁰. Libraries were amplified with 7-10 cycles using NEBNext High-Fidelity 2X PCR Master-Mix (NEB, #M0541S). To remove primer dimers and DNA fragments shorter than 150bp, libraries were purified with SpeedBeads™ beads. 1.8X volume of SpeedBeads™ beads (81µl beads for 45µl PCR reaction) was used to clean PCR amplified DNA library. The quality of the DNA libraries was assessed using Agilent HS DNA Screen Tape (Agilent, Santa Clara, CA). The concentration of the DNA libraries was measured by Qubit Fluorometric Quantitation. Equimolar DNA libraries were pooled to a final concentration of 2nM. The ATAC-seq libraries were sequenced by Novogene (Novogene, UK) on Illumina HiSeq-2500 using pair-ends PE150 in biological duplicates (Illumina, San Diego, CA).

Sequencing Data Processing and Bioinformatics

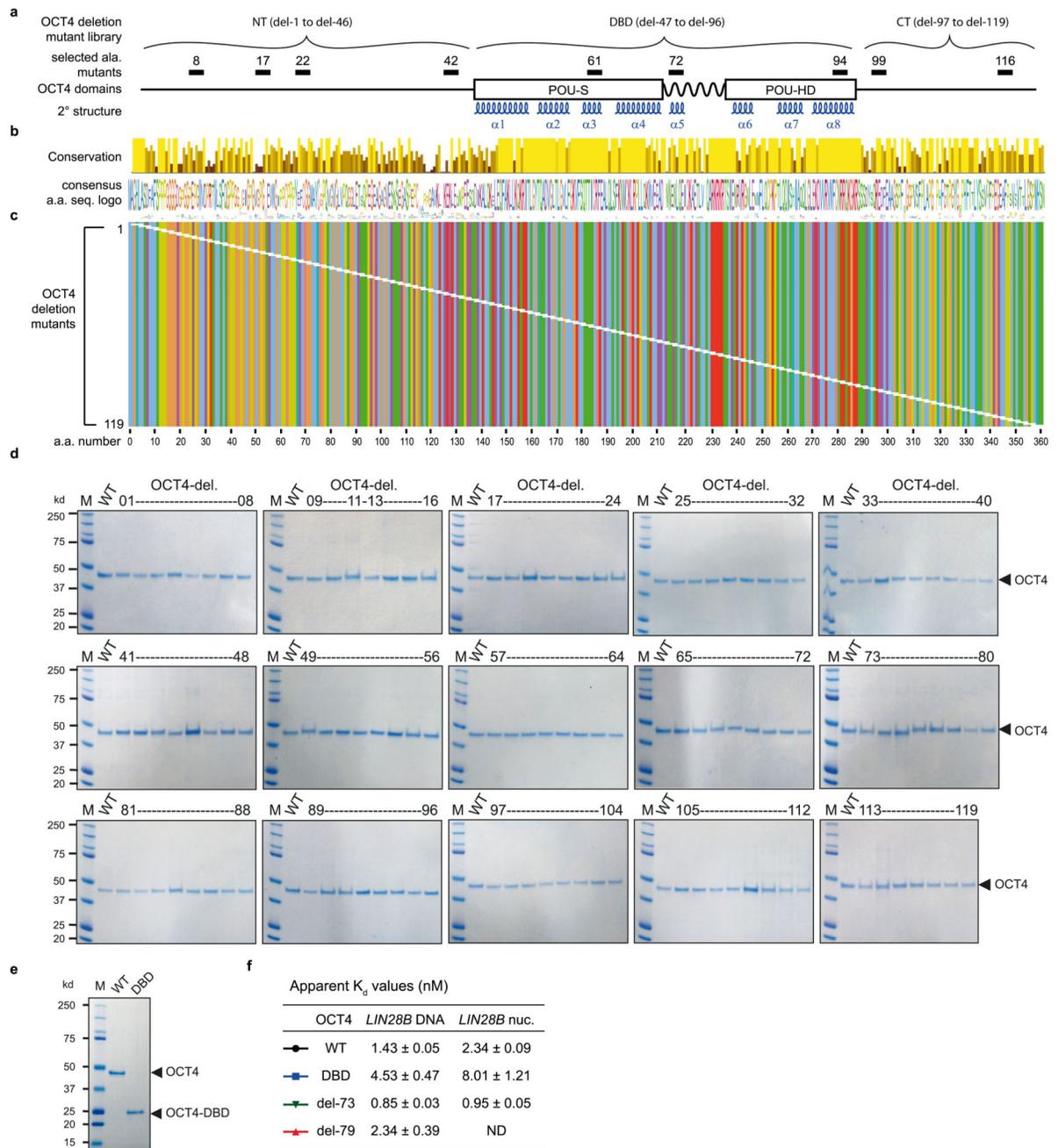
All sequencing data analysis are provided in the Supplementary protocols.

Statistics and Reproducibility

The binding assays of the OCT4-WT or mutants were carried out at least three times and representative EMSA gels were presented. All reprogramming and pluripotency

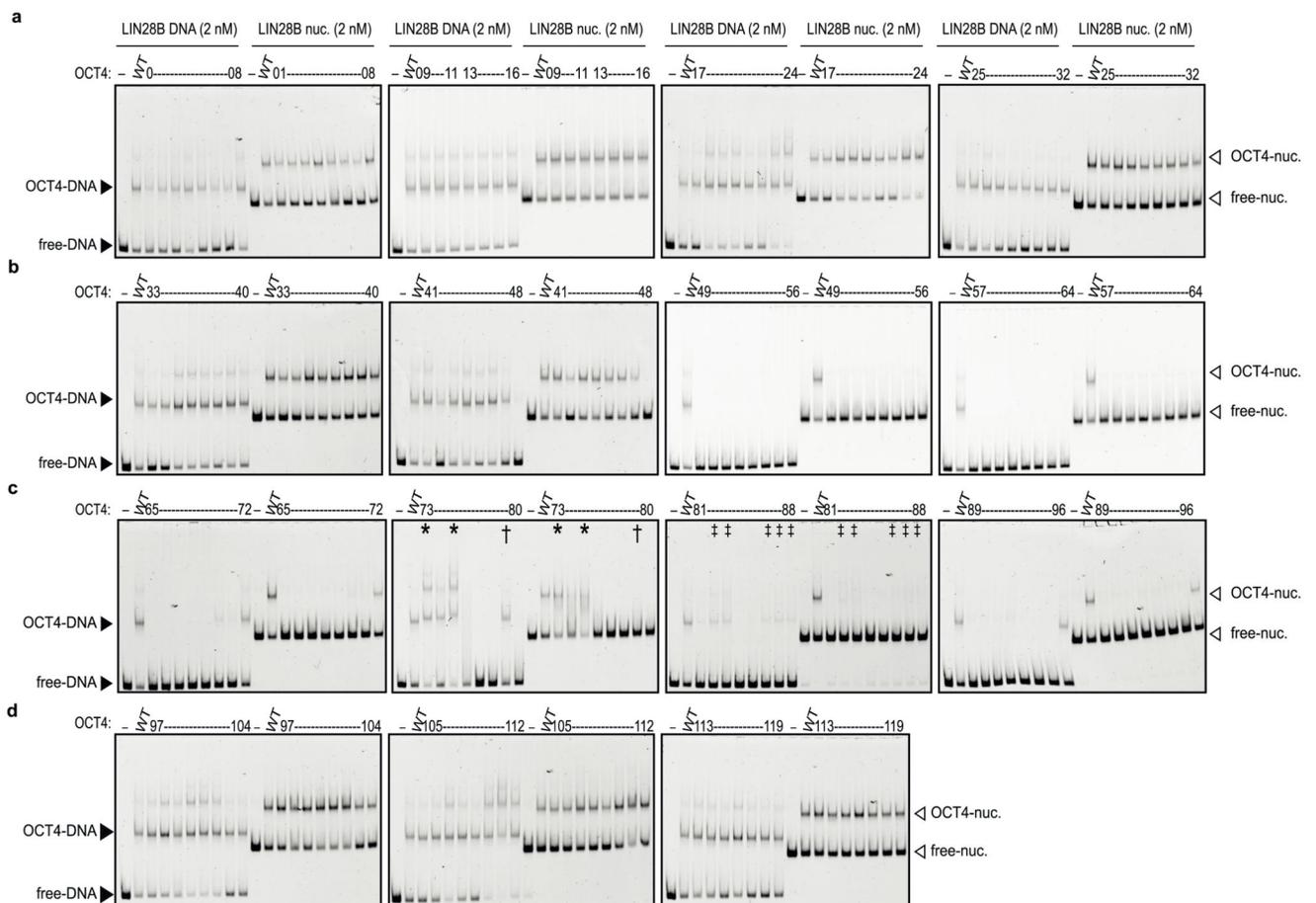
maintenance experiments were independently repeated at least three times with similar results obtained. Data were plotted and analysed using GraphPad Prism 7 (GraphPad Software) with built-in tests and non-linear fitting algorithm as indicated in the figure legends and the Methods section. OCT4 ChIP-seq was carried out from three independent ChIP experiments pooled together as indicated in the Methods. ATAC-seq data were generated from two independent experiments, which were then merged post sequencing for further analysis. Statistical tests and P values from ChIP-seq, ATAC-seq and MNase-seq data were obtained as indicated in the Methods. Number of images and particles used for the *LIN28B*-nucleosome cryo-EM structure determination are summarized in extended data figure 4a and explained in the Methods and Supplementary methods.

Extended Data

**Extended Data Fig. 1. Mutagenesis strategy to define the sub-domains of OCT4**

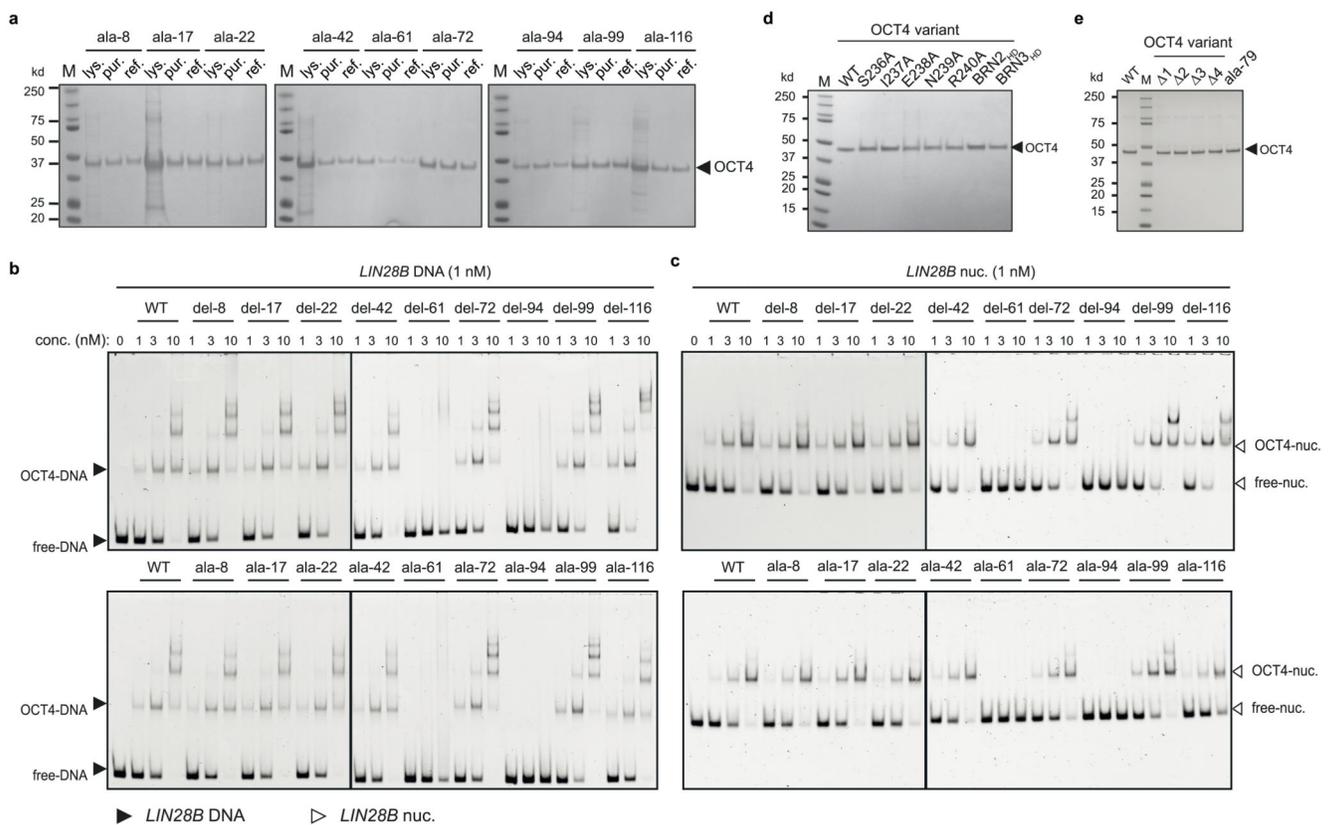
a, Schematic representation of OCT4 protein, indicating OCT4-DBD in white boxes and the linker region in black zigzag line. The α -helices (α 1-8) defining the protein secondary structure are indicated by blue spirals. NT and CT regions are indicated by black lines. The deletion mutations spanning OCT4 domains are indicated above the curly brackets. Selected 5X alanine-stretch substitution are indicated by black boxes and numbered according to

the equivalent deletion counterparts. **b**, The conservation plot (top), and the amino-acid consensus sequence (bottom) from multiple sequence alignment of OCT4 orthologs across mammals. **c**, Multiple sequence alignment of the 119 OCT4 derivatives, highlighting the deleted five a.a. regions as white boxes. The a.a. sequence is colour-coded following the colour scheme of the consensus sequence in (b). **d**, The generation of 118 out of 119 recombinant OCT4 deletion derivatives (del-1 to del-119) from bacteria as analyzed by SDS-PAGE and Coomassie staining. The respective OCT4 bands run at the expected size when compared to the sizes of the protein standards (M). The concentration of each OCT4 deletion derivative was measured from the intensity of the respective Coomassie-stained band relative to that of OCT4-WT run on the same gel. Apart from OCT4-WT, del-73, and del-79, all OCT4 recombinants were generated from one purification experiment. **e**, SDS-PAGE analysis showing Coomassie blue stained bands representing the recombinant OCT4-DBD with similar purity and concentration to the full-length counterpart. At least three biological replicates were used. **f**, The apparent dissociation constant (Kd) OCT4-WT and mutants were determined from the binding curves in (Fig. 1d). ND indicates that Kd values were not determined due to non-saturated binding. Errors represent to the goodness of the fit to the experimental data.



Extended Data Fig. 2. The affinity of OCT4 deletion mutants for naked DNA versus nucleosomes.

a-d, EMSA gels showing the affinity of each recombinant OCT4 deletion mutant (118 in total) for Cy5-labelled LIN28B-DNA (left half, black arrowheads) and LIN28B-nucleosome (right half, white arrowheads). No protein (-) and wildtype recombinant OCT4 (WT) were used as controls for each EMSA. The binding was measured at equilibrium using 2nM of LIN28B-DNA or LIN28B-nucleosome incubated with 3nM of each OCT4 deletion derivative. The number of OCT4 deletions are indicated above each EMSA. **c**, Asterisks indicate del-73 and del-75 mutants showing enhanced affinity for DNA and nucleosomes. Dagger indicates del-79 mutant that can bind DNA but not nucleosomes. Double dagger indicates del-82-83 and del-86-87 OCT4 mutants with partial nucleosome uncoupling as the interaction of these mutants with DNA is also diminished. Apart from OCT4-WT, del-73, del-75 and del-79, which were repeated at least three times, the EMSA screen was carried out once.



Extended Data Fig. 3. OCT4 alanine-stretch substitution and deletion mutants display similar affinity for DNA and nucleosomes

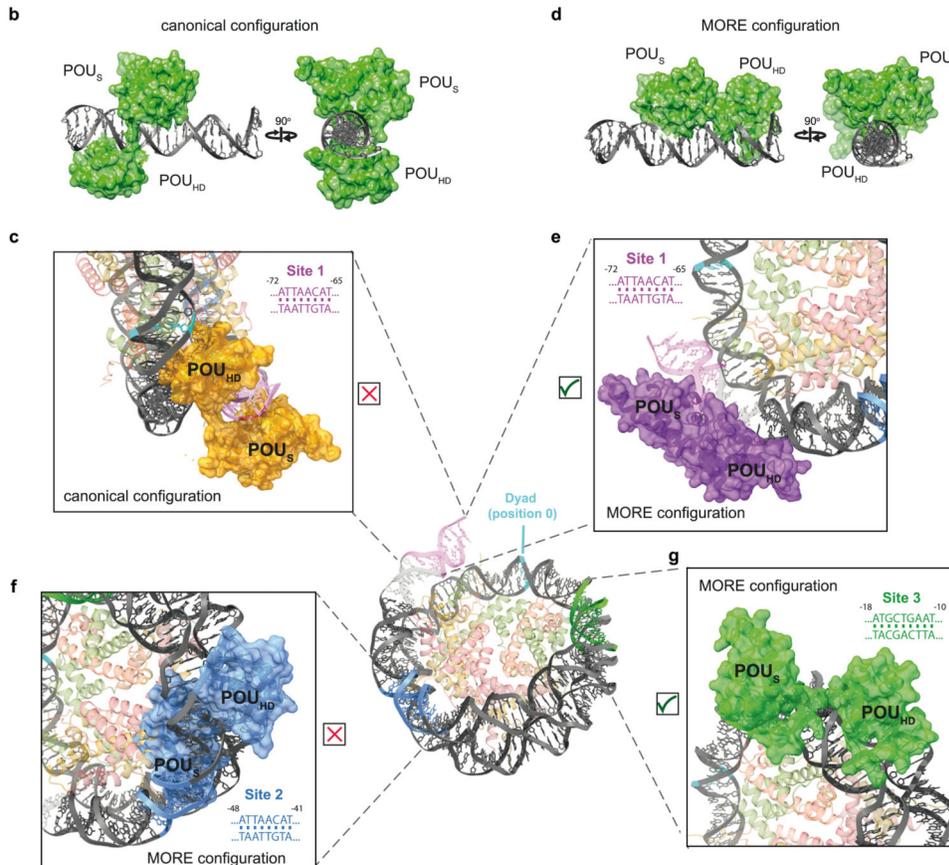
a, SDS-PAGE analysis showing Coomassie blue stained bands representing the selected OCT4 alanine stretch mutants shown in (Extended data figure 1a). The fractions from the lysates (lys.), purified under denaturing conditions (pur.), and refolded (ref.) are shown for each mutant. The respective OCT4 bands (arrow) run at the expected size when compared to the sizes of the protein standards (M). **b-c**, EMSA showing the affinity of increasing amounts of recombinant deletion (top panels) and the equivalent alanine substitution (bottom

panels) OCT4 mutants for Cy5-labelled LIN28B-DNA (b, black arrowheads) and LIN28B-nucleosome (c, white arrowheads). OCT4 concentrations (nM) are indicated above each lane. **d-e**, same as (a) for OCT4 alanine point-mutants within (a.a. 236-240) and BRN2HD, BRN3HD hybrids (d), and delta 1-4 and ala-79 OCT4 mutants (e). The concentration of each OCT4 mutant was measured from the intensity of the respective Coomassie-stained band relative to that of OCT4-WT run on the same gel.

a-e are representative images of at least three biological replicates.

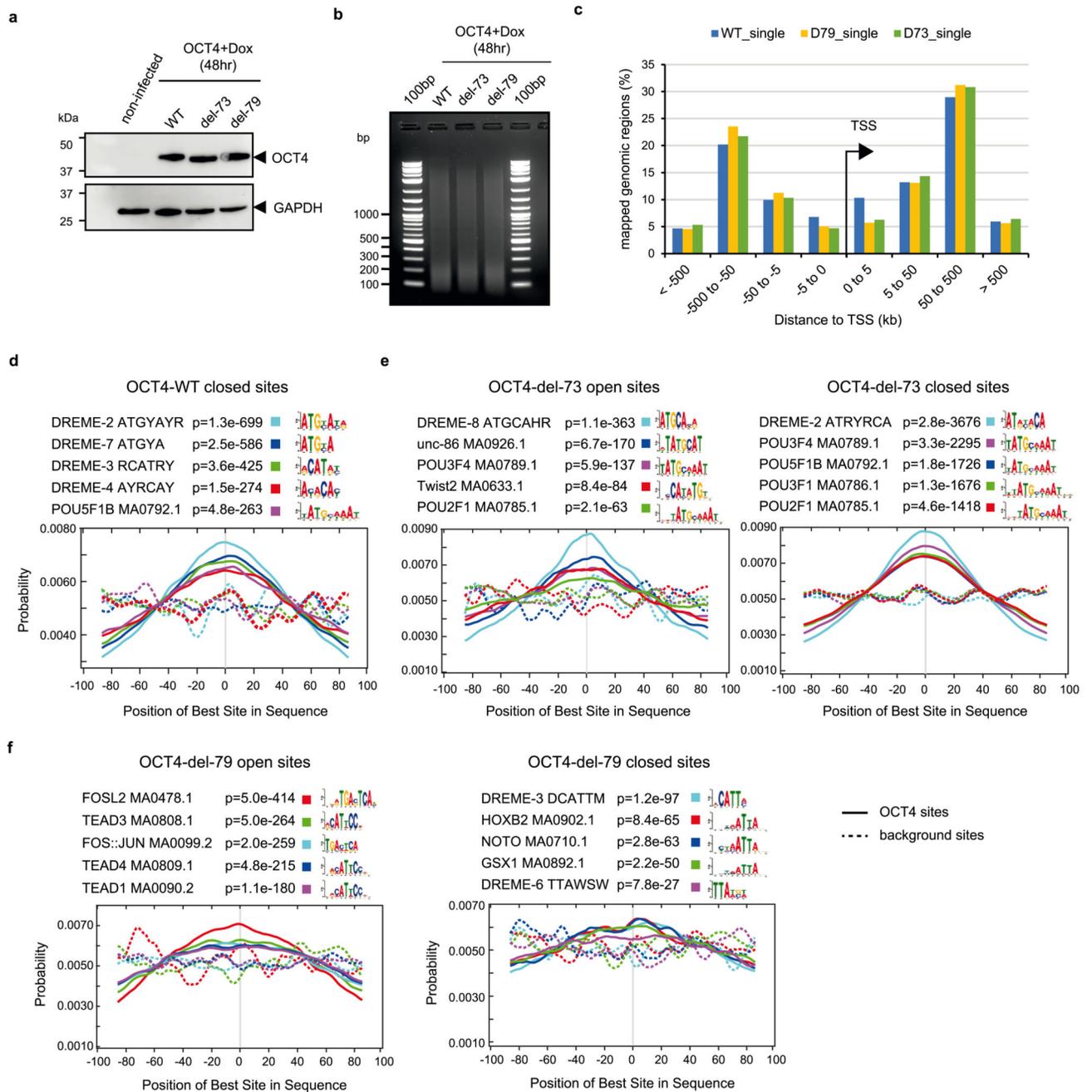
a

Data collection and processing		Refinement		mean <i>B</i> factors (Å ²)	
Magnification	130 000	Initial model (PDB code)	6IPU	Protein	32.57
Voltage (kV)	300	Model resolution (Å)	3.3	Ligand	83.01
Electron exposure (e ⁻ Å ⁻²)	37.8	FSC threshold	0.143	R.M.S. deviations	
Defocus range (µm)	-1.5 to -3.3	Map sharpening <i>B</i> factor (Å ²)	0	Bond lengths (Å)	0.006
Pixel size (Å)	0.65	Model composition		Bond angles (°)	0.695
Symmetry imposed	no	Non-hydrogen atoms	11488	Validation	
Initial particle images (no.)	4013899	Protein residues	769	MolProbity score	1.90
Final particle images (no.)	217170	Nucleotide	262	Clashscore	11.96
Map resolution (Å)	3.5	Ligands	0	Poor rotamers (%)	0
FSC threshold	0.143	Water	0	Ramachandran plot	
				Favoured (%)	95.62
				Allowed (%)	4.38
				Disallowed (%)	0



Extended Data Fig. 4. OCT4 configurations compatible with binding to the LIN28B-nucleosome

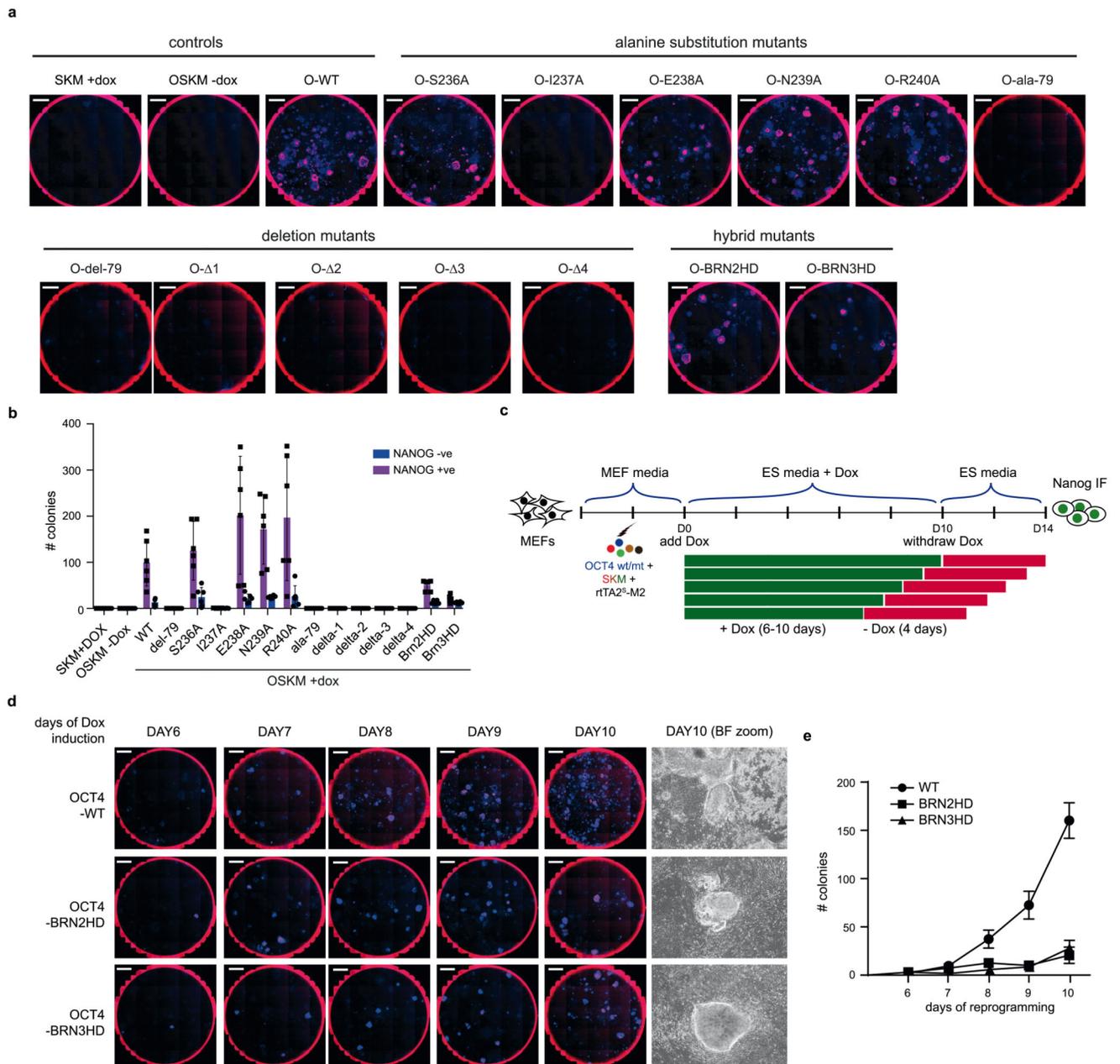
a, Table summarizing cryo-EM imaging data collection and analysis to resolve the 3D structure of the LIN28B-nucleosome. **b and d**, Surface-filled representations of OCT4-DBD (green) bound to DNA (cartoon representation in grey) in the canonical (b) and MORE configuration (d). **c**, Surface-filled model of OCT4-DBD (orange) bound in the canonical configuration to the LIN28B-nucleosome at site 1 (pink), indicating incompatible binding due to steric clashes. **e-g**, Surface-filled models of OCT4-DBD in the MORE configuration bound to LIN28B-nucleosome at site 1 (e, magenta), site 2 (g, blue) and site 3 (f, green) to demonstrate that sites 1 and 3 are compatible (green ticks) with OCT4 nucleosome binding. Whereas OCT4 is prevented (red cross) from binding the LIN28B-nucleosome at site 2 due to steric clashes.



Extended Data Fig. 5. OCT4 nucleosome binding enables specific access to closed chromatin

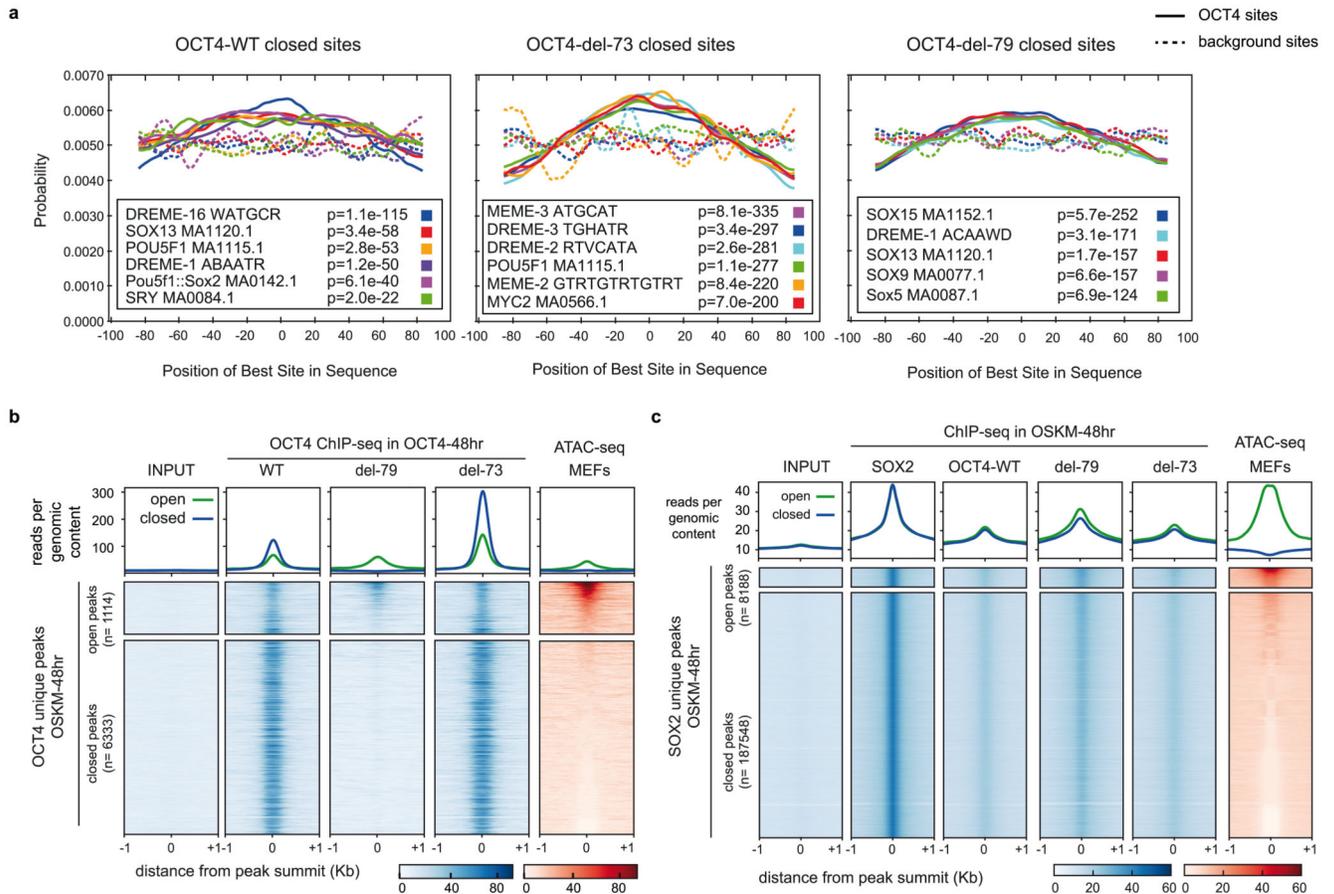
a, Western blot analysis showing antibody specificity and protein levels of OCT4-WT, del-73 and del-79 derivatives expressed in MEFs after lentiviral infection and Dox induction for 48hr. GAPDH was used as a loading control. Protein sizes are indicated. Non-infected MEFs were used as a negative control. Data representative of 3 biological replicates. **b**, Agarose gel electrophoresis showing equivalent chromatin fragmentation used for ChIP-seq of OCT4-WT and mutants. Data representative of 3 biological replicates. **c**, Bar plots showing the distribution of OCT4 WT and mutants peaks relative to TSS as measured by

GREAT analysis. **d-f**, The central enrichment of *de novo* motifs identified by DREME and TOMTOM analysis as measured by CentriMO analysis within the OCT4-WT closed sites (d), OCT4-del-73 open and closed sites (e), and the OCT4-del-79 open and closed sites (f). No motifs were identified within OCT4-WT open sites. Each motif enrichment is color coded (solid line) and controlled against background sequences (dotted line) that are 1 kb away from each OCT4 peak. A significant central enrichment is measured by CentriMO (p value) as indicated. Motif logos are indicated on top of each plot. Statistical analyses are provided in source data for extended data fig. 5.



Extended Data Fig. 6. Pioneer function of OCT4 is essential for inducing pluripotency

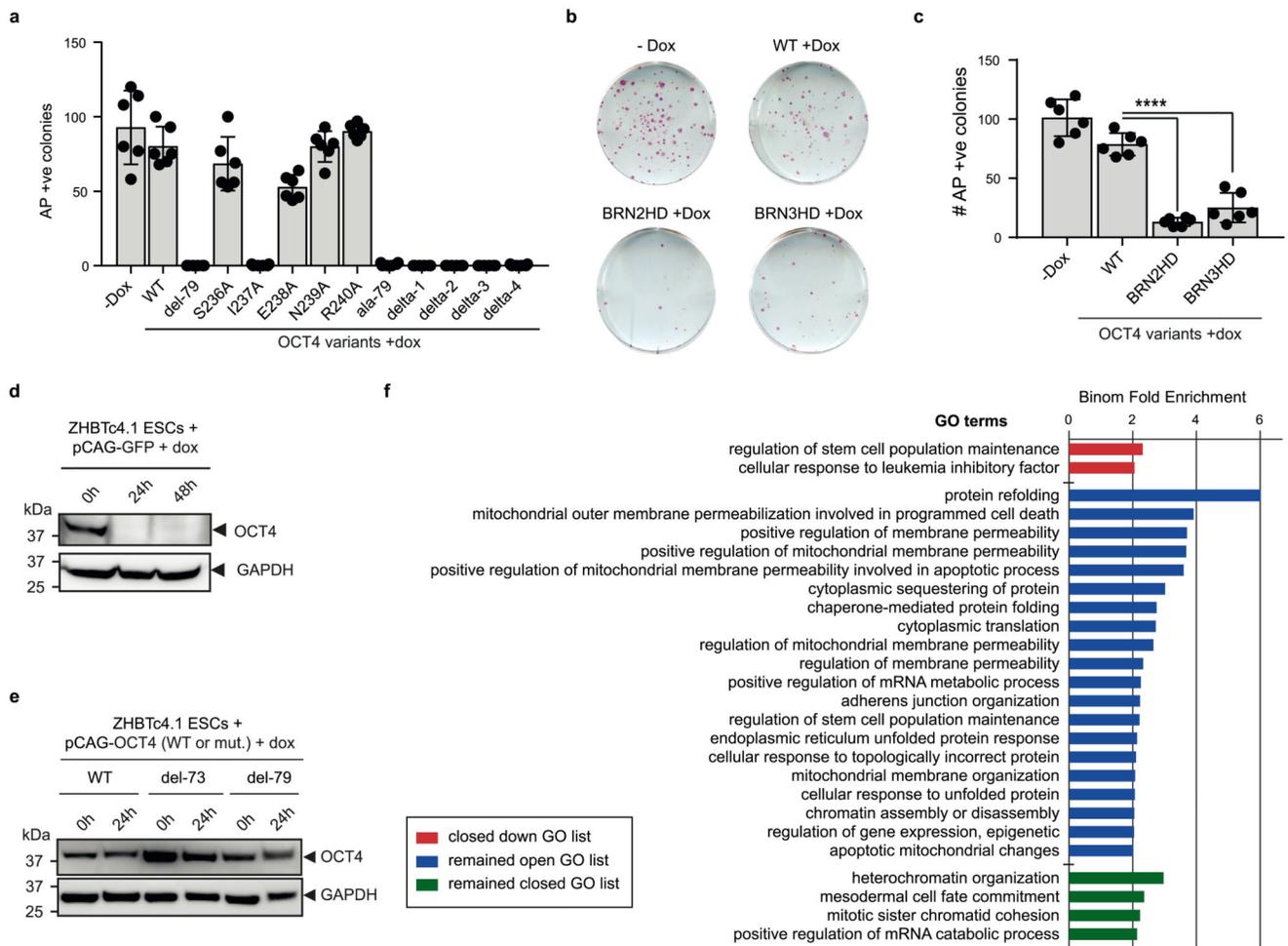
a, Whole-well images showing DAPI fluorescence (blue), Nanog immuno-fluorescence (red) and merged (magenta) on day 16 (4 days after Dox withdrawal) of reprogramming MEFs to iPSCs using OCT4 WT and mutants. MEFs infected with SKM (no OCT4) or OSKM in the absence of Dox (-Dox) were used as negative controls. 5 mm scale bars. Images are representative of 6 biological replicates. **b**, Bar plots showing the efficiency of reprogramming to pluripotency as quantified by the number of Nanog positive colonies (purple bars) versus DAPI-only colonies (Nanog -ve, blue bars) for OCT4-WT and mutants, each in combination with SKM. SKM +Dox and OSKM +/- Dox were used as controls (bars 1-3). Averages of n=6 independent biological replicates are shown (error bars indicate \pm s.d.). Data points are represented in squares (Nanog +ve) and circles (Nanog -ve). **c**, Schematic showing the experimental workflow of reprogramming MEFs to iPSCs under increasing durations of ectopic OSKM expression using Dox. **d**, Representative whole-well images of 3 biological replicates showing DAPI fluorescence (blue), Nanog immuno-fluorescence (red) and merged (magenta) of MEFs ectopically expressing OCT4-WT, BRN2HD or BRN3HD hybrid mutants along with SKM for 6, 7, 8, 9, and 10 days. Dox-independent iPSC colonies were counted 4 days after Dox withdrawal. Representative bright field images of Dox-independent colonies (10 days induction) are shown in the left-hand panels. 5 mm scale bars. **e**, Line plots showing the reprogramming efficiency of OCT4-WT (circles), BRN2HD (squares) and BRN3HD (triangles) hybrid mutants during the different stages explained in (c and d). Averages of n=3 independent biological replicates are shown (error bars indicate \pm s.d.).



Extended Data Fig. 7. SOX2 can drive the pioneer deficient OCT4 to closed chromatin

a, The distribution of *de novo* motifs identified by DRME and TOMTOM analysis as measured by CentriMO analysis within closed sites targeted by OCT4-WT, OCT4-del-73, and OCT4-del-79. The significance of central enrichment is measured by CentriMO (p value) as indicated. Statistical analysis are provided in source data extended data fig. 7. **b**, Read-density heatmaps showing the ChIP-seq enrichment of OCT4 WT, del-73, and del-79 mutants when expressed alone in MEFs (blue) and MEFs ATAC-seq signal (red) spanning \pm 1kb from the centre of OCT4 unique peaks identified in early reprogramming (OSKM-48h). The analyzed sequences are rank-ordered from high to low according to the ATAC-seq reads within the central 300 bp of each peak. The ATAC-enriched (open) sequences are separated from the ATAC-depleted (closed) sequences and the percentage of each category is indicated. The total number of peaks (n) for each OCT4 derivative are shown to the right. The colour scales are indicated below. Input DNA is also shown. Line plots to show the average ChIP-seq or ATAC-seq enrichment are above each corresponding heatmap. **c**, Read-density heatmaps showing the ChIP-seq enrichment of SOX2, OCT4 WT, del-73, and del-79 mutants in early reprogramming (blue) and MEFs ATAC-seq signal (red) spanning \pm 1kb from the centre of SOX2 unique sites identified in OSKM-48h. The analyzed sequences are rank-ordered from high to low according to the ATAC-seq reads within the central 300 bp of each peak. The ATAC-enriched (open) sequences are separated from the ATAC-depleted (closed) sequences and the percentage of each category is indicated. The total number of

peaks (n) and the colour scales are indicated. Input DNA is also shown. Line plots to show the average ChIP-seq or ATAC-seq enrichment are above each corresponding heatmap.



Extended Data Fig. 8. The conserved pioneer activity of OCT4 is crucial for maintaining pluripotency

a, Bar plots showing the rescued AP positive ESC colonies in the presence of Dox and OCT4-WT or a del-79-derived mutant. Averages of $n=6$ independent biological replicates are shown (error bars indicate \pm s.d.). **b**, Whole well images showing AP positive ZHBTc4.1 ESCs in the absence of Dox (-Dox) or the presence of Dox (+Dox) either ectopically expressing OCT4 WT or one of the POUHD hybrids. Images are representative of 6 biological replicates. **c**, Bar plots showing the number of AP positive colonies in the presence of Dox for OCT4 WT, BRN2HD and BRN3HD as in (b). POUHD hybrids failed to rescue pluripotency compared to OCT4 WT (compare WT bars to BRN2HD and BRN3HD). Two-tailed t-test was used to assess statistical significance. **** indicates $P < 0.0001$. Averages of $n=6$ independent biological replicates are shown (error bars indicate \pm s.d.) as represented in (b). **d**, Western blot analysis showing the diminished levels of Oct4 after treating ZHBTc4.1 ESCs, ectopically expressing GFP, with Dox for 24h and 48h. GAPDH was used as loading control. Images are representative of 2 biological replicates.

e, Representative Western blots from 2 biological replicates showing ZHBTc4.1 ESCs, ectopically expressing OCT4-WT or mutants after 24h of Dox treatment. f, GO analysis of genes targeted by OCT4 at sites that closed down (red bars), remained open (blue bars) or remained closed (green bars) after treating ZHBTc4.1 ESCs, expressing ectopic OCT4-WT or mutants, with Dox for 24h.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Profs D. O'Carroll, I. Chambers, W. Bickmore, and K. Zaret for comments on the manuscript and insightful discussions. We thank Prof S. Rosser, the Edinburgh Genome Foundry and the UK Centre for Mammalian Synthetic Biology for support with generating synthetic DNA constructs. We thank Edinburgh Genomics for the sequencing service. We thank Sukanija Thillainadarasan for assistance with making the OCT4 alanine-stretch mutant constructs. We thank Dr Marcus Wilson for kindly providing the tailless H3, H4, and the acidic patch histone mutants and help with nucleosome reconstitution. We thank Natasha Lukoyanova for assistance with cryo-EM data collection. The work was supported by an MRC career development award (MR/N024028/1) to A.S that supported A.S, B.O and E.H-P. The UK Research Councils' Synthetic Biology for Growth programme and of the BBSRC, EPSRC and MRC (BB/M0180404/1) supported the work of G.A.R. The early detection committee CRUK-OHSU Project Award (C65925/A26986) to A.S supported G.A.R. A Ph.D scholarship from the Darwin Trust of Edinburgh supported M.R.O. The BBSRC-EASTBIO Ph.D training program (BB/M010996/1) supported I.G. An MRC career development award (MR/R008795/1) to P.J.R. supported the work of P.J.R and M.S. Cryo-EM data for this investigation were collected at Birkbeck College, University of London with financial support from the Wellcome Trust (202679/Z/16/Z and 206166/Z/17/Z).

Data Availability

The cryo-EM map of the *LIN28B* nucleosome core particle have been deposited in the Electron Microscopy Data Bank with database code EMD-12453. The atomic coordinates of the *LIN28B* NCP model have been deposited in the Protein Data Bank with PDB ID: 7NL0. The human nucleosome core particle (PDB: 6IPU)⁵¹ and the OCT4-DBD-DNA crystal structures (PDB:3L1P)⁵² were used for model building. All next generation sequencing data generated as part of this study have been deposited in the Gene Expression Omnibus (GEO) under one super-series with the accession number GSE168142, which is composed of the sub-series GSE167632 and GSE168141. Previously published ATAC-seq and MNase-seq data that were re-analysed here are available under accession codes GSE98124³⁵ and GSM1004653¹³, respectively. Data were aligned to the *Mus musculus* genome assembly MGScv37 (mm9) with accession: PRJNA20689. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

References

1. Iwafuchi-Doi M, Zaret KS. Pioneer transcription factors in cell reprogramming. *Genes Dev.* 2014; 28 :2679–2692. [PubMed: 25512556]
2. King HW, Klose RJ. The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. *eLife.* 2017; 6 e22631 [PubMed: 28287392]
3. Gaspar-Maia A, Alajem A, Meshorer E, Ramalho-Santos M. Open chromatin in pluripotency and reprogramming. *Nature Reviews Molecular Cell Biology.* 2010; 12 :36.

4. Chen X, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*. 2008; 133 :1106–1117. [PubMed: 18555785]
5. Soufi A, Donahue G, Zaret KS. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell*. 2012; 151 :994–1004. [PubMed: 23159369]
6. Knaupp AS, et al. Transient and Permanent Reconfiguration of Chromatin and Transcription Factor Occupancy Drive Reprogramming. *Cell Stem Cell*. 2017; 21 :834–845. e836 [PubMed: 29220667]
7. Li D, et al. Chromatin Accessibility Dynamics during iPSC Reprogramming. *Cell Stem Cell*. 2017; 21 :819–833. e816 [PubMed: 29220666]
8. Zviran A, et al. Deterministic Somatic Cell Reprogramming Involves Continuous Transcriptional Changes Governed by Myc and Epigenetic-Driven Modules. *Cell Stem Cell*. 2019; 24 :328–341. e329 [PubMed: 30554962]
9. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006; 126 :663–676. [PubMed: 16904174]
10. Chronis C, et al. Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell*. 2017; 168 :442–459. e420 [PubMed: 28111071]
11. Chen J, et al. Hierarchical Oct4 Binding in Concert with Primed Epigenetic Rearrangements during Somatic Cell Reprogramming. *Cell Reports*. 2016; 14 :1540–1554. [PubMed: 26832419]
12. Soufi A, et al. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell*. 2015; 161 :555–568. [PubMed: 25892221]
13. Teif VB, et al. Genome-wide nucleosome positioning during embryonic stem cell development. *Nature structural & molecular biology*. 2012; 19 :1185–1192.
14. Cirillo LA, et al. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell*. 2002; 9 :279–289. [PubMed: 11864602]
15. Cirillo LA, et al. Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J*. 1998; 17 :244–254. [PubMed: 9427758]
16. Cirillo LA, Zaret KS. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell*. 1999; 4 :961–969. [PubMed: 10635321]
17. Iwafuchi-Doi M, et al. The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. *Mol Cell*. 2016; 62 :79–91. [PubMed: 27058788]
18. Zhu F, et al. The interaction landscape between transcription factors and the nucleosome. *Nature*. 2018; 562 :76–81. [PubMed: 30250250]
19. Fernandez Garcia M, et al. Structural Features of Transcription Factors Associating with Nucleosome Binding. *Mol Cell*. 2019; 75 :921–932. e926 [PubMed: 31303471]
20. Iwafuchi M, et al. Gene network transitions in embryos depend upon interactions between a pioneer transcription factor and core histones. *Nature Genetics*. 2020; 52 :418–427. [PubMed: 32203463]
21. Michael AK, et al. Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science*. 2020 eabb0074
22. Tapia N, et al. Dissecting the role of distinct OCT4-SOX2 heterodimer configurations in pluripotency. *Sci Rep*. 2015; 5 13533 [PubMed: 26314899]
23. Nichols J, et al. Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. *Cell*. 1998; 95 :379–391. [PubMed: 9814708]
24. Ambrosetti DC, Basilico C, Dailey L. 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. *Molecular and Cellular Biology*. 1997; 17 :6321–6329. [PubMed: 9343393]
25. Avilion AA, et al. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev*. 2003; 17 :126–140. [PubMed: 12514105]
26. Remenyi A, et al. Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev*. 2003; 17 :2048–2059. [PubMed: 12923055]

27. Boyer LA, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*. 2005; 122 :947–956. [PubMed: 16153702]
28. Zaret KS, Carroll JS. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev*. 2011; 25 :2227–2241. [PubMed: 22056668]
29. Kong X, et al. Functional interplay between the RK motif and linker segment dictates Oct4-DNA recognition. *Nucleic Acids Research*. 2015; 43 :4381–4392. [PubMed: 25870414]
30. Echigoya K, et al. Nucleosome binding by the pioneer transcription factor OCT4. *Scientific Reports*. 2020; 10 11832 [PubMed: 32678275]
31. Jerabek S, et al. Changing POU dimerization preferences converts Oct6 into a pluripotency inducer. *EMBO reports*. 2017; 18 :319–333. [PubMed: 28007765]
32. Huertas J, MacCarthy CM, Schöler HR, Cojocaru V. Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding. *Biophysical Journal*. 2020; 118 :2280–2296. [PubMed: 32027821]
33. Zhou K, Gaullier G, Luger K. Nucleosome structure and dynamics are coming of age. *Nature Structural & Molecular Biology*. 2019; 26 :3–13.
34. von Hippel P. Protein-DNA recognition: new perspectives and underlying themes. *Science*. 1994; 263 :769–770. [PubMed: 8303292]
35. Benchetrit H, et al. Direct Induction of the Three Pre-implantation Blastocyst Cell Types from Fibroblasts. *Cell Stem Cell*. 2019; 24 :983–994. e987 [PubMed: 31031139]
36. Niwa H, Masui S, Chambers I, Smith AG, Miyazaki J. Phenotypic complementation establishes requirements for specific POU domain and generic transactivation function of Oct-3/4 in embryonic stem cells. *Mol Cell Biol*. 2002; 22 :1526–1536. [PubMed: 11839818]
37. Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet*. 2000; 24 :372–376. [PubMed: 10742100]
38. Smith AG, et al. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature*. 1988; 336 :688–690. [PubMed: 3143917]
39. Blomquist P, Li Q, Wrangé ö. The Affinity of Nuclear Factor 1 for Its DNA Site Is Drastically Reduced by Nucleosome Organization Irrespective of Its Rotational or Translational Position. *Journal of Biological Chemistry*. 1996; 271 :153–159.
40. Taylor IC, Workman JL, Schuetz TJ, Kingston RE. Facilitated binding of GAL4 and heat shock factor to nucleosomal templates: differential function of DNA-binding domains. *Genes & Development*. 1991; 5 :1285–1298. [PubMed: 2065977]
41. Mueller B, et al. Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction. *Genes Dev*. 2017; 31 :451–462. [PubMed: 28356342]
42. Mieczkowski J, et al. MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. *Nat Commun*. 2016; 7 11485 [PubMed: 27151365]
43. Voong LN, et al. Insights into Nucleosome Organization in Mouse Embryonic Stem Cells through Chemical Mapping. *Cell*. 2016; 167 :1555–1570. e1515 [PubMed: 27889238]
44. Iurlaro M, et al. Mammalian SWI/SNF continuously restores local accessibility to chromatin. *Nature Genetics*. 2021; 53 :279–287. [PubMed: 33558757]
45. Schick S, et al. Acute BAF perturbation causes immediate changes in chromatin accessibility. *Nature Genetics*. 2021; 53 :269–278. [PubMed: 33558760]
46. Hockemeyer D, et al. A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. *Cell Stem Cell*. 2008; 3 :346–353. [PubMed: 18786421]
47. Kim J, et al. An iPSC line from human pancreatic ductal adenocarcinoma undergoes early to invasive stages of pancreatic cancer progression. *Cell Rep*. 2013; 3 :2088–2099. [PubMed: 23791528]
48. Dyer PN, et al. Reconstitution of Nucleosome Core Particles from Recombinant Histones and DNA. *Methods in Enzymology*. 2003; 375 :23–44.
49. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*. 2012; 9 :671–675. [PubMed: 22930834]

50. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013; 10 :1213–1218. [PubMed: 24097267]
51. Sharma D, et al. PARP1 exhibits enhanced association and catalytic efficiency with γ H2A.X-nucleosome. *Nature Communications*. 2019; 10 5751
52. Esch D, et al. A unique Oct4 interface is crucial for reprogramming to pluripotency. *Nat Cell Biol*. 2013; 15 :295–301. [PubMed: 23376973]

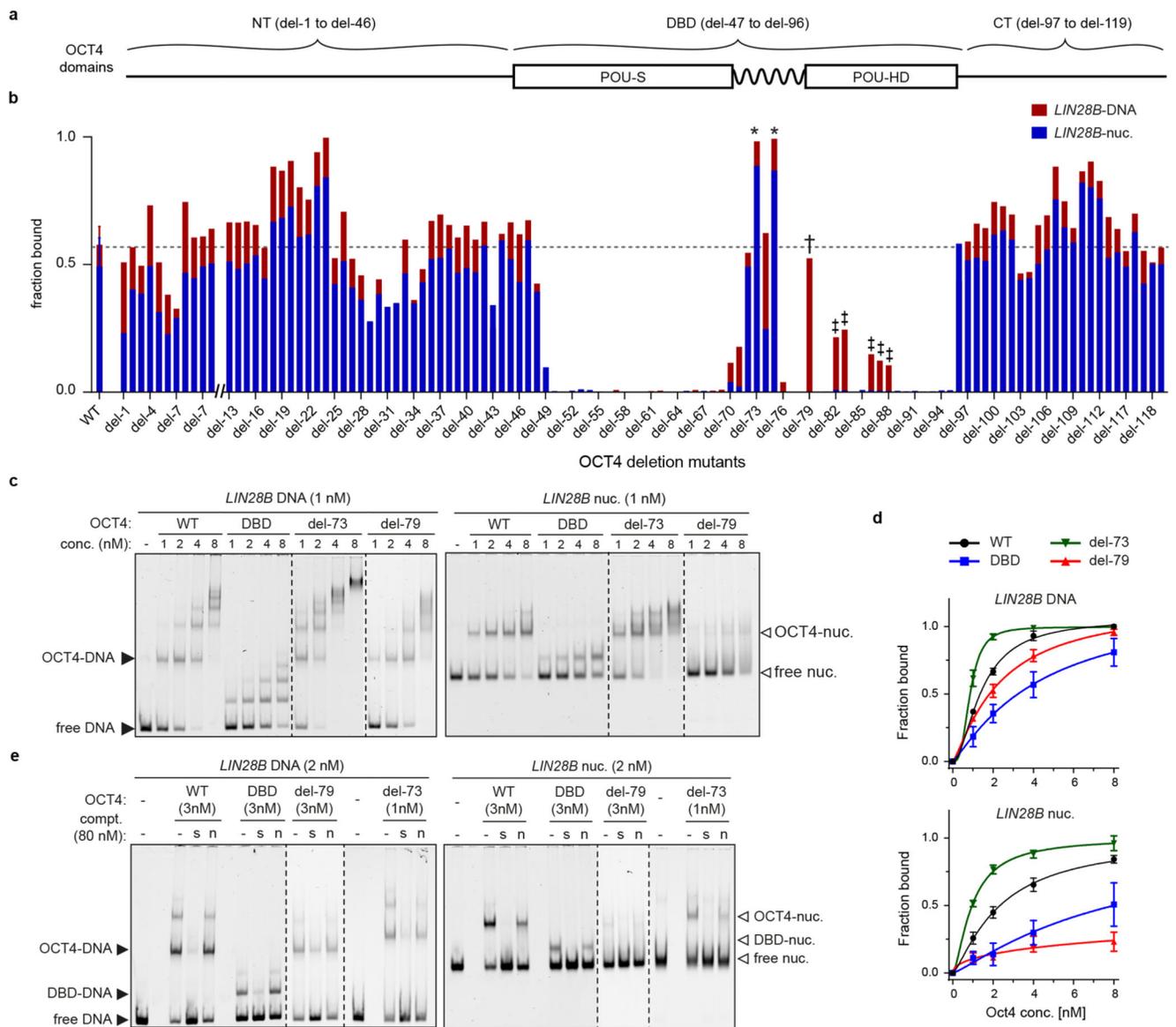


Fig. 1. Nucleosome binding is encoded by OCT4-DBD and yet can be uncoupled from DNA binding.

a, A schematic representation of OCT4 domain organisation, indicating the mutants generated within each domain. **b**, Bar chart showing the affinity of each OCT4 variant for *LIN28B*-DNA (red) versus *LIN28B*-nucleosome (blue). Asterisks indicate mutants showing enhanced affinity for DNA and nucleosomes. Dagger indicates del-79 mutant that can bind DNA but not nucleosomes. Double dagger indicates del-82-83 and del-86-87 OCT4 mutants with partial nucleosome uncoupling as the interaction of these mutants with DNA is also diminished. Dashed line represent average affinity of OCT4-WT for DNA. **c**, Representative EMSA from 3 biological repeats showing the affinity of increasing amounts of recombinant OCT4-WT, DBD, del-73 and del-79 for Cy5-labelled *LIN28B*-DNA (black arrowheads) and *LIN28B*-nucleosome (white arrowheads). Concentration range is indicated above each EMSA. Higher order bands are also observed at elevated OCT4 concentrations,

corresponding to multiple OCT4 molecules bound per one *LIN28B*-DNA at different sites. **d**, Binding curves of OCT4 WT and mutants to naked DNA (top) and nucleosomes (bottom) were generated by fitting the Michaelis–Menten equation to the experimental bound DNA or nucleosomes as a function of OCT4 concentration [nM] from n=3 independent biological replicates represented in (c). Data are presented as mean values +/- SD. **e**, Representative EMSA from 3 biological repeats, showing the affinity of OCT4-WT and mutants for Cy5-labelled *LIN28B*-DNA (black arrowheads) and *LIN28B*-nucleosome (white arrowheads) in the presence of 40-fold molar excess of non-labelled specific competitor (“s” lanes) or non-specific competitor (“n” lanes) or absence of competitor (“-” lanes).

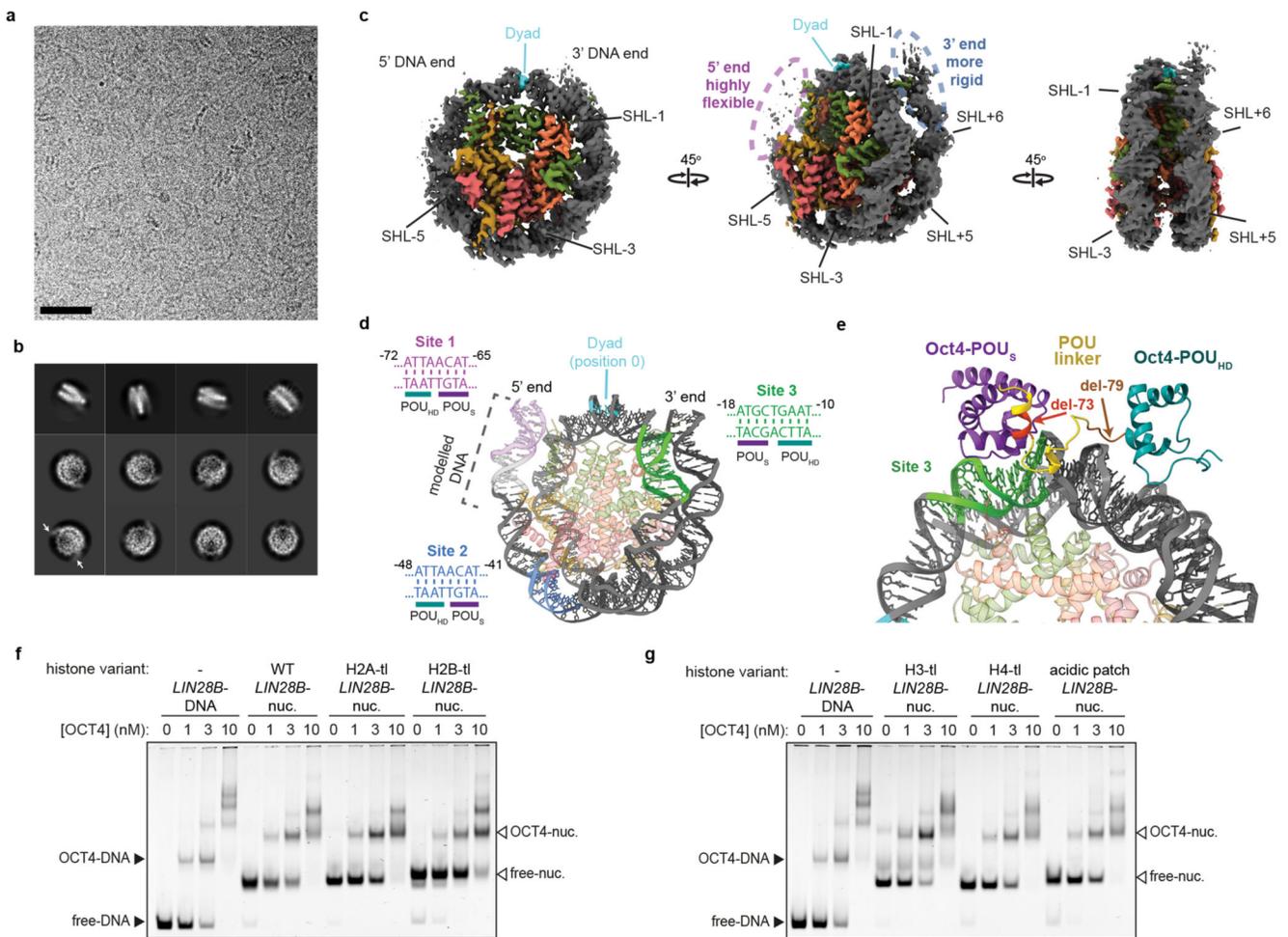


Fig. 2. Structural insights into OCT4 interaction with the LIN28B-nucleosome.

a, Representative cryo-EM micrograph of the LIN28B-nucleosome. 30nm scale bar. A summary of data collection and processing is shown in extended data figure 4a. A total of 12892 images (movies) were collected, of which 12232 images were used for particle picking. **b**, 2D-class averages showing the LIN28B-nucleosome in various orientations. Arrows indicate DNA projecting from the nucleosome core. A total number of 4,013,899 single particles were picked, of which 2,848,487 particles were used for the final 2D classification. **c**, Surface representation of the LIN28B-nucleosome cryo-EM map. Notable features of the LIN28B-DNA (grey) such the dyad (cyan) and entry/exit sites are indicated. The 5' and 3' DNA ends are circled to indicate flexibility. Histones H2A, H2B, H3 and H4 are coloured gold, pink, green and orange, respectively. **d**, The position of three OCT4 binding sites (site 1: Magenta, site 2: Blue, site 3: Green) are shown including details of the orientation of the DNA motifs recognized by POU_S and POU_{HD}. **e**, Cartoon representation of OCT4-DBD in the MORE configuration bound to LIN28B-nucleosome on site 3 with the linker separating the POU domains (yellow) and residues deleted in the del-73 (red) and del-79 (brown) mutants indicated. **f**, **g**, Representative EMSA from 3 biological replicates showing the affinity of OCT4 for various nucleosome mutants containing either a histone

tailless H2A-tl, H2B-tl, H3-tl, H4-tl mutant or an acidic-patch H2A-H2B histone mutant as compared to wild-type histones (WT) or naked DNA.

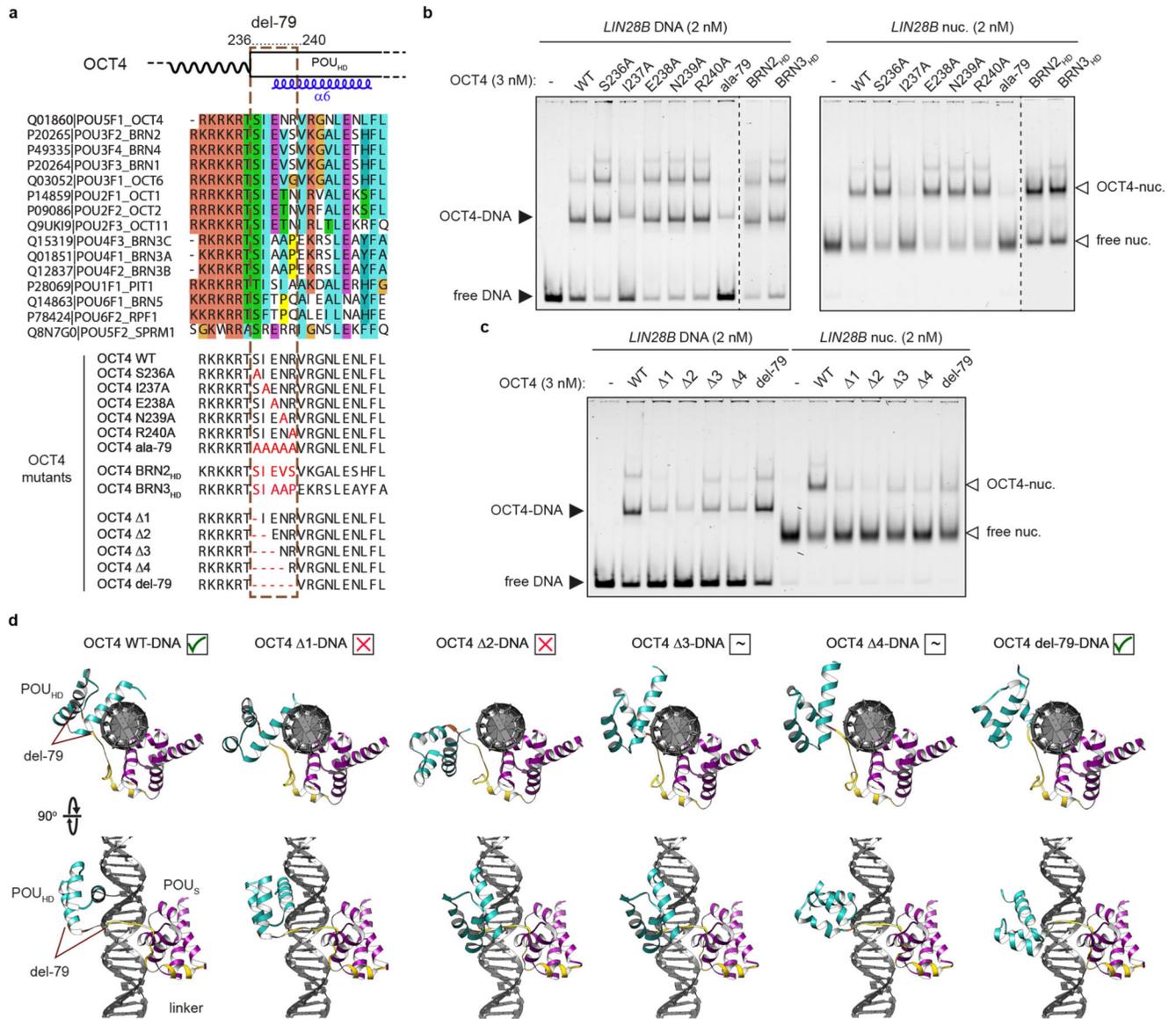


Fig. 3. Restricting the adaptation of OCT4 to nucleosomes uncouples DNA-binding.

a, Multiple sequence alignment showing part of the homeodomain that contains the OCT4-del-79 region (dashed box) from various TFs belonging to the human POU family. The mutated residues within OCT4-del-79 are indicated in red. Name of proteins and mutants are listed on the left and a schematic showing the protein organisation is shown above. **b**, **c**, Representative EMSA from 3 biological replicates showing the affinity of each recombinant OCT4-WT, alanine-point mutants and POU_{HD} hybrids (**b**) and OCT4- mutants (**c**) for Cy5-labelled *LIN28B*-DNA (left half, black arrowheads) and *LIN28B*-nucleosome (right half, white arrowheads). **d**, Cartoon representations showing the top (top panels) and side views (bottom panels) of homology models of OCT4-DBD-WT and -mutants in complex with DNA, highlighting the POU_S (Magenta), POU_{HD} (Cyan), linker (Yellow) and del-79 region (Brown). Ticks, crosses, and tilde symbols to represent strong, none or weak DNA-binding, respectively.

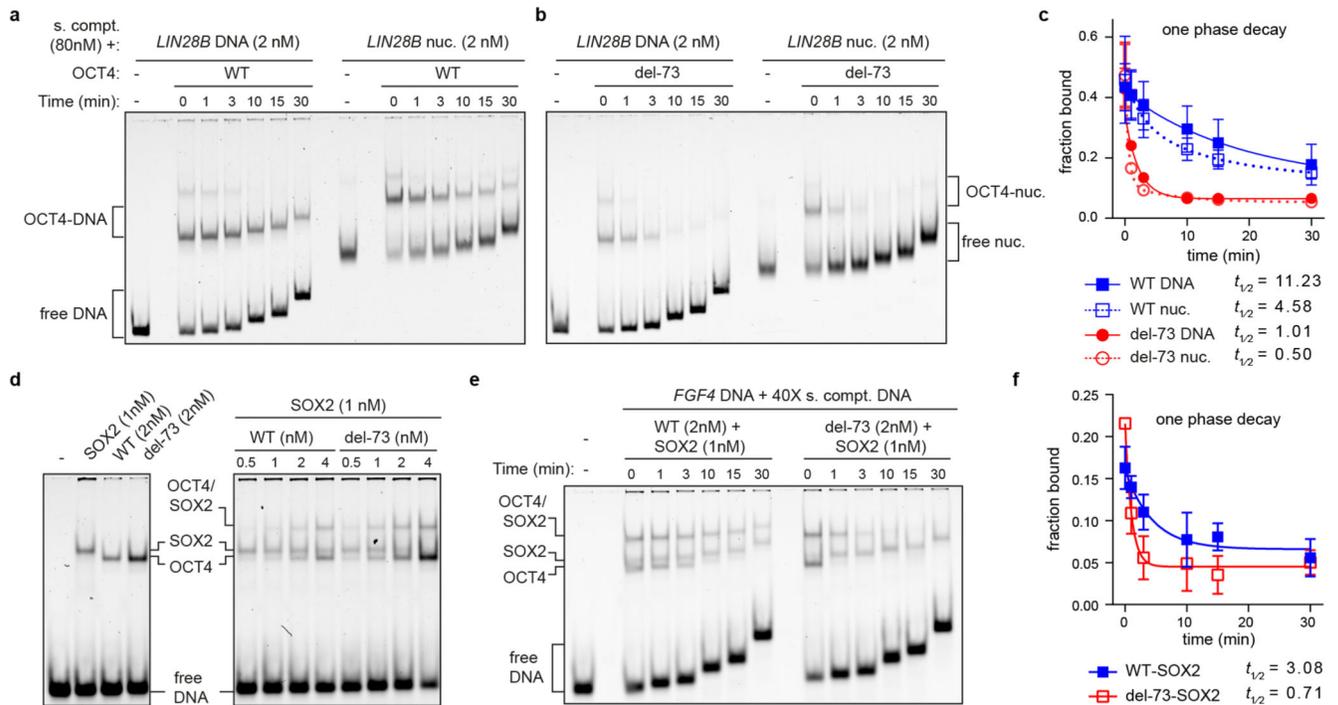


Fig. 4. Accelerated DNA and nucleosome binding impedes OCT4-SOX2 cooperativity.

a, Representative EMSA from 3 biological replicates showing the decay of OCT4-WT in complex with Cy5-labelled *LIN28B*-DNA (left-hand side) and nucleosomes (right-hand side) after adding excessive amounts of specific competitor (non-labelled) for increasing time intervals (indicated above each lane) after equilibrium. **b**, same as (a) for OCT4-del73. **c**, Binding decay curves to determine the half-life ($t_{1/2}$) of OCT4-WT (blue) and del-73 (red) in complex with *LIN28B*-DNA (solid line) or *LIN28B*-nucleosome (dotted line), which were fitted to experimental data of OCT4-WT (squares) and del-73 (circles). Data are presented as mean values \pm SD from $n=3$ biological replicates represented in (a-b). The obtained half-life values are indicated below. **d**, Representative EMSA from 3 biological replicates showing OCT4-WT, del-73, and SOX2 binding individually to Cy5-labelled *FGF4* DNA (left panel), and the cooperative interaction of OCT4-WT or del-73 with SOX2 under increasing amounts of OCT4 and constant amounts SOX2 (right panel). The corresponding OCT4 and SOX2-DNA complexes are indicated. **e**, Representative EMSA from 3 biological replicates showing the decay of OCT4-WT-SOX2 (left) or OCT4-del-73-SOX2 (right) in complex with Cy5-labelled *FGF4*-DNA after adding excessive amounts of specific competitor (non-labelled) for increasing time intervals (indicated above each lane) after equilibrium. **f**, Binding decay curves to determine the half-life ($t_{1/2}$) of OCT4-WT-SOX2 (blue) and OCT4-del-73-SOX2 (red) in complex with *FGF4*-DNA, which were fitted to experimental data of OCT4-WT (filled squares) and del-73 (empty squares). Data are presented as mean values \pm SD from $n=3$ biological replicates represented in (e). The obtained half-life values are indicated below.

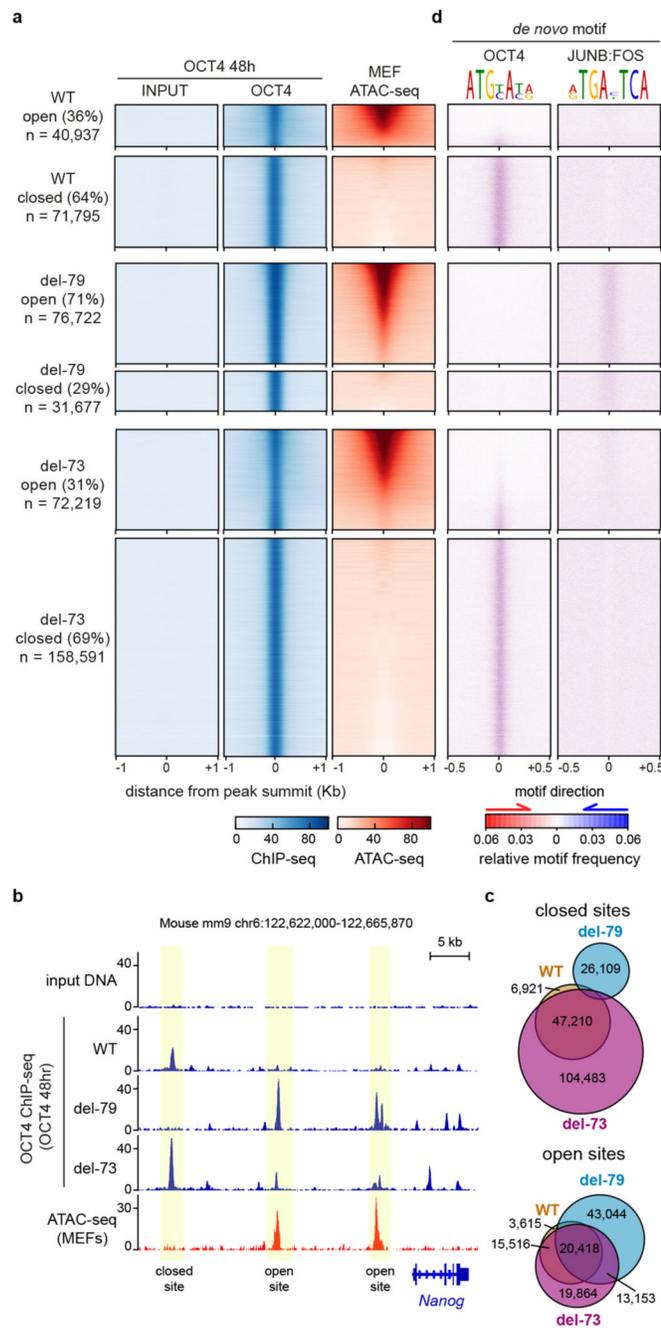


Fig. 5. Nucleosome binding enables OCT4 to target specific sites within inaccessible chromatin. **a**, Read-density heatmaps showing normalized ChIP-seq intensity of OCT4 WT, del-del-73, and del-79 mutants overexpressed in MEFs for 48h (blue) compared to input DNA (blue) and ATAC-seq signal in MEFs (red) spanning ± 1 kb from OCT4 peak summits. The analysed sequences are rank-ordered based on ATAC-seq read density within the central 300 bp of OCT4 peaks. The ATAC-enriched (open) sequences are separated from the ATAC-depleted (closed) sequences and the percentage of the sites within each category are indicated. Total number of peaks (n) for each OCT4 derivative are shown to the right.

The colour scale is indicated at the bottom. **b**, Genome browser screenshot showing input DNA and the ChIP-seq profiles for each OCT4-WT and mutants and ATAC-seq in MEFs within the *Nanog* locus indicated above. **c**, Venn diagram to display the extent of binding overlap between OCT4-WT and mutants within closed (left) and open sites (right). **d**, The enrichment of motifs discovered by *de novo* analysis (motif logos on top) on sites targeted by OCT4-WT and mutants overexpressed in MEFs for 48h. Heatmaps showing motif scores on the top (red) and bottom (blue) DNA strands spanning across ± 500 bp from each OCT4 peak summit. The colour scale of motif scores along with motif direction are indicated at the bottom. The sites are rank-ordered based on ATAC-seq shown in (a).

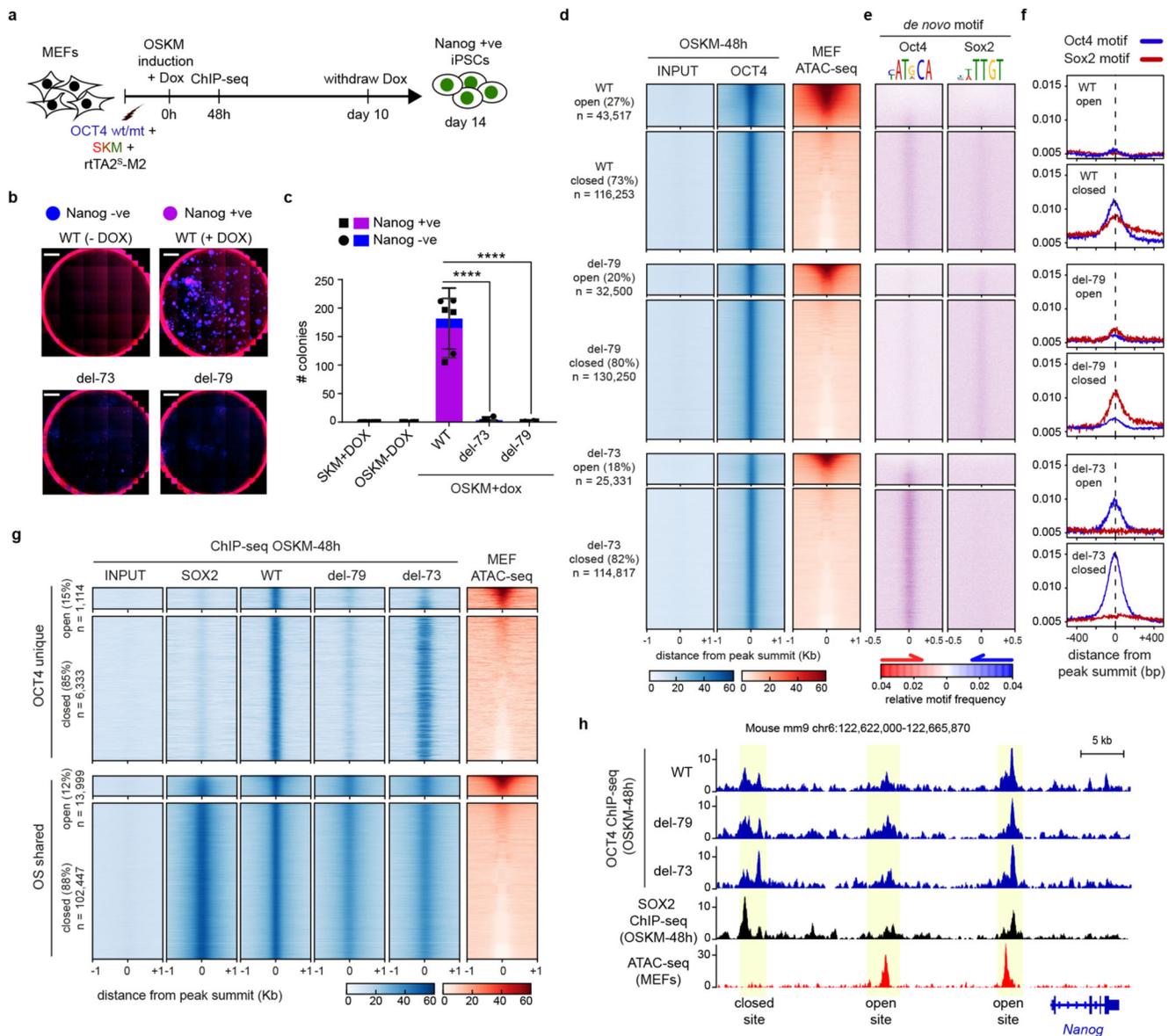


Fig. 6. The pioneer activity of OCT4 is required for reprogramming.

a, Schematic showing the experimental workflow for reprogramming MEFs to iPSCs.

b, Representative whole-well images from 3 biological replicates showing DAPI staining (blue), Nanog immuno-fluorescence (red) and merged (magenta) colonies after reprogramming MEFs to iPSCs using OCT4-WT or mutants. **c**, Bar plots showing the efficiency of iPSCs reprogramming using OCT4-WT and mutants. Nanog positive colonies (purple bars) versus DAPI-only colonies (Nanog -ve, blue bars) for each. SKM +Dox and OSKM -/+ Dox were used as controls (bars 1-3). Averages of n=3 independent biological replicates are shown (error bars indicate +/- SD). Dunnett's multiple comparisons test (two-tailed) was used to assess statistical significance. **** indicates P < 0.0001. 5 mm scale bars are indicated. **d**, Read-density heatmaps showing the ChIP-seq intensity of OCT4-WT, and mutants in OSKM48h (blue) and MEFs ATAC-seq signal (red) spanning ± 1kb around

OCT4 peak summits. Sequences are rank-ordered according to the ATAC-seq and separated as open or closed. The total number of peaks (n) and the colour scales are indicated. Input DNA is also shown. **e**, Heatmaps showing the enrichment of OCT4 and SOX2 motifs discovered by *de novo* motif analysis (motif logos on top) across the sites targeted by OCT4-WT and mutants in OSKM48h and rank ordered by ATAC-seq as in (d). Motifs are scored on the Top (red) and bottom (blue) DNA strands across ± 500 bp from each OCT4 peak summit. The colour scale of motif scores along with motif direction are indicated. **f**, Line plots quantifying the average enrichment of OCT4 motif (blue) and SOX2 motif (red) shown in (e). **g**, Read-density heatmaps showing ChIP-seq for SOX2, OCT4-WT, and mutants in OSKM48h (blue) and MEFs ATAC-seq (red) spanning ± 1 kb around the centre of OCT4 unique sites (top panels) and OCT4-SOX2 shared sites (bottom panels). The analysed sequences are rank-ordered according to the ATAC-seq and separated as open or closed. **h**, Genome browser screenshot showing the ChIP-seq profiles for OCT4-WT and mutants as well as SOX2 in OSKM48hr and ATAC-seq in MEFs within the Nanog locus.

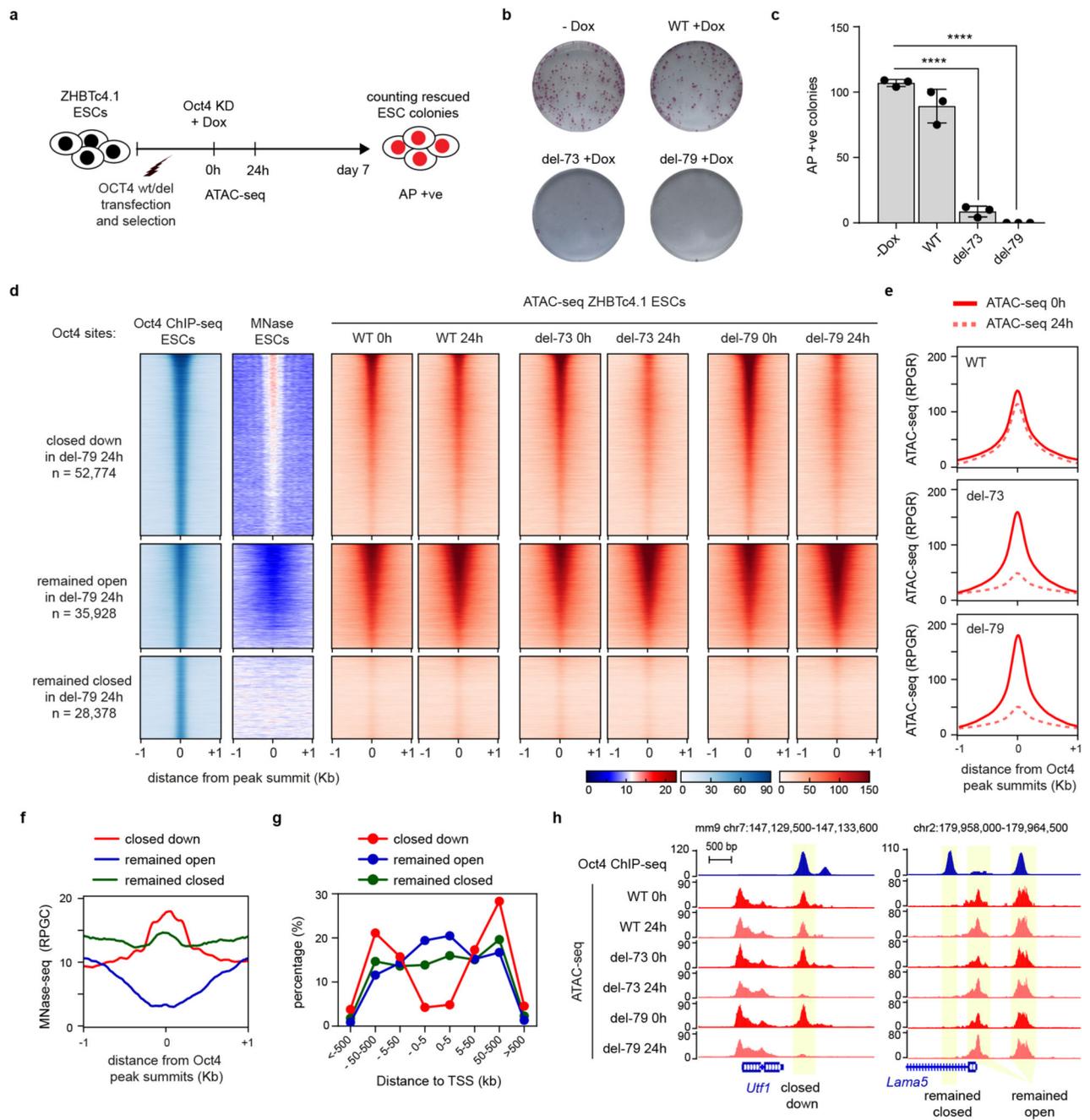


Fig. 7. OCT4 pioneer activity is essential for maintaining open enhancers in ESCs.

a, Schematic diagram depicting the experimental setup for maintaining ZHBTc4.1 ESCs undifferentiated with ectopic OCT4. **b**, Whole well images showing alkaline phosphatase positive (AP +ve) ZHBTc4.1 ESCs in the absence of Dox (-Dox) or the presence of Dox (+Dox) along with the ectopic OCT4-WT or mutants. Representative of 3 biological replicates. **c**, Bar plots showing the number of AP positive colonies in the presence of Dox for OCT4 WT and mutants relative to that in the absence of Dox (-Dox). Averages of n=3 independent biological replicates are shown (error bars indicate +/- SD) as represented

in (b). Dunnett's multiple comparisons test (two-tailed) was used to assess statistical significance. **** indicates $P < 0.0001$. **d**, Read-density heatmaps showing normalized ChIP-seq intensity of endogenous Oct4 in ESCs (blue), MNase-seq intensity (blue-to-red) and ATAC-seq signal (red) spanning ± 1 kb from the centre of Oct4 peaks. ATAC-seq before (0h) and after (24h) endogenous Oct4 knocked-down in ZHBTc4.1 ESCs and rescued with OCT4-WT or mutants is shown. The Oct4-enriched sequences were grouped based on whether they lost accessibility (closed down), maintained accessibility (remained open), or stayed inaccessible (remained closed) upon Oct4-KD in ESCs transfected with Oct4-del-79. The total number of OCT4 peaks (n) for each group and the colour scales are indicated. **e**, Line plots showing average ATAC-seq signal of sites that closed down in (d). **f**, Line plots showing average MNase-seq signal of sites that closed down (red), remained open (blue), or remained closed in (green) as shown in (d). **g**, The distance to TSS distribution of sites that closed down (red), remained open (blue), or remained closed (green). **h**, Genome browser tracks showing Oct4 ChIP-seq in relation to ATAC-seq before (0h, red) and after (24h, light red) Oct4 knock-down in ZHBTc4.1 ESCs transfected with OCT4-WT or mutants. Representative peaks are highlighted in yellow.