HHS Public Access

Author manuscript

Mol Cell. Author manuscript; available in PMC 2024 June 15.

Published in final edited form as:

Mol Cell. 2023 June 15; 83(12): 1970–1982.e6. doi:10.1016/j.molcel.2023.05.030.

Structural mechanism of *LIN28B* nucleosome targeting by OCT4

Ruifang Guan[1,2], **Tengfei Lian**[1,2], **Bing-Rui Zhou**[1], **David Wheeler**[1], **Yawen Bai**[1,3,*]

[1]Laboratory of Biochemistry and Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

[2]These authors equally contributed to this work

[3]Lead Contact

SUMMARY

Pioneer transcription factors are essential for cell fate changes by targeting closed chromatin. OCT4 is a crucial pioneer factor that can induce cell reprogramming. However, the structural basis of how pioneer factors recognize the *in vivo* nucleosomal DNA targets is unknown. Here, we determine the high-resolution structures of the nucleosome containing human *LIN28B* DNA and its complexes with the OCT4 DNA binding region. Three OCT4s bind the pre-positioned nucleosome by recognizing non-canonical DNA sequences. Two use their POUS domains while the other uses the POUS-loop-POUHD region; POUHD serves as a wedge to unwrap ~25 base pair DNA. Our analysis of previous genomic data and determination of the *ESRRB*-nucleosome-OCT4 structure confirmed the generality of these structural features. Moreover, biochemical studies suggest that multiple OCT4s cooperatively open the H1-condensed nucleosome array containing the *LIN28B* nucleosome. Thus, our study suggests a mechanism of how OCT4 can target the nucleosome and open closed chromatin.

eTOC blurb

Guan et al. determined the structures of the nucleosome containing a genomic DNA and its complexes with OCT4. They find that multiple OCT4 molecules bind the nucleosome and its two DNA binding domains work together to unwrap the nucleosomal DNA. The study suggest how OCT4 may open closed chromatin locally.

Graphical Abstract

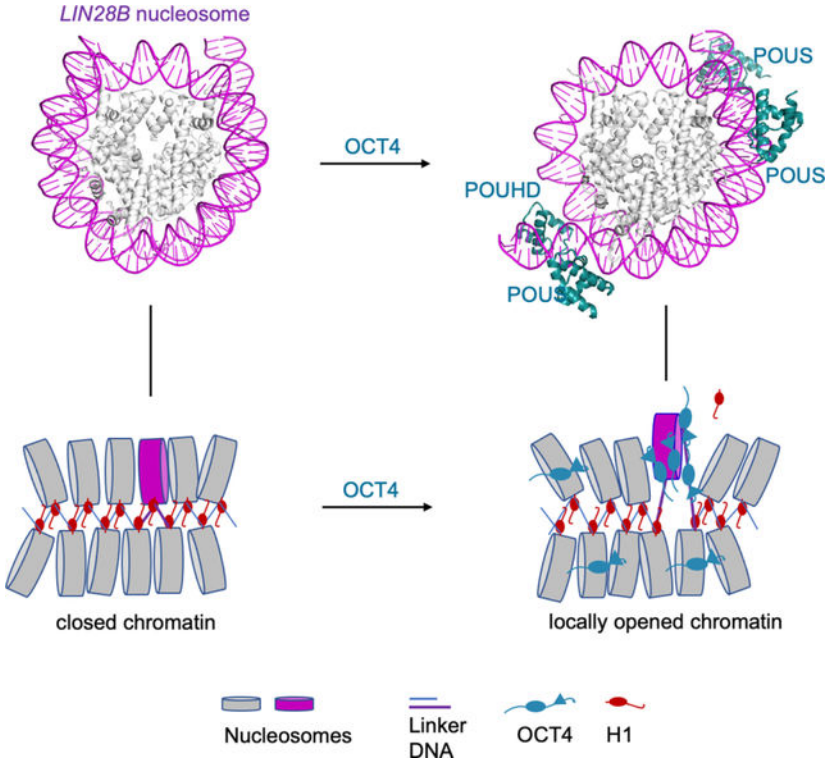
*Correspondence: baiyaw@mail.nih.gov.

Author contributions

Y.B. conceived the project. R.G. and T.L. conducted the biochemical experiments and solved the structure. B-R. Z. provided the scFv and H1.4 proteins and reconstituted the nucleosome array. D.W. analyzed the ChIP-seq and MNase-seq data. Y.B., R.G., and T.L. analyzed the structure and wrote the paper.

Declaration of interests

Authors declare no competing interests.



INTRODUCTION

Gene regulation plays a crucial role in cell fate control. However, in eukaryotic cells, genomic DNA is packaged into chromatin by associating core and linker histones to form nucleosomes and chromatosomes.[1,2] Together with the compaction of chromatin to higher-order structures, a significant fraction of the DNA surface is not accessible for transcription factors. Nevertheless, a special group of transcription factors, termed pioneer factors, can recognize the closed chromatin sites (DNase I resistant) at enhancers.[3,4] They facilitate the subsequent recruitment of other transcription factors to regulate gene expression and play critical roles in cell differentiation and development.[5, 6] For example, the forced expression of transcription factors OCT4, SOX2, KLF4, and c-MYC can induce pluripotent stem cells from somatic cells by accessing closed chromatin sites.[7,8,9] Functional genomics and biochemical studies reveal that OCT4 binds the nucleosome containing the human enhancer *LIN28B* DNA (162 bp, Figure 1A) during the reprogramming of fibroblasts to pluripotent cells.[9–12] The *LIN28B* locus is important for pluripotency reprogramming and OCT4 binding to the *LIN28B* nucleosome precedes *LIN28B* gene activation, which is silent in human fibroblasts and remains silent after 48-hour induction of OCT4, SOX2, KLF4, and c-MYC.[8,9]

To understand how OCT4 recognizes the *LIN28B* nucleosome, we want to solve the structure of the nucleosome bound to OCT4. However, determining the high-resolution structures of the nucleosomes containing genomic DNA sequences and their complexes with the pioneer factors is technically challenging. Genomic nucleosomes are fragile and

pioneer factors tend to dissociate during sample preparation for structural studies by single particle cryo-electron microscopy (cryo-EM). So far, our knowledge of the interactions between pioneer factors and nucleosomal DNA comes from studies using model systems with non-genomic DNA sequences that bind the core histones tightly. These earlier studies have led to various binding modes of pioneer transcription factors. [13,14,15,16,17,18] In the case of OCT4, it only uses the POUS domain to recognize its canonical DNA motif (ATGC) that is incorporated into the Widom 601 (W601) DNA near the entry-exit site of the nucleosome.[15] Other biochemical and modeling studies suggest that OCT4 may also use the POUS domain to recognize the canonical motif in the inner region of the nucleosomal DNA.[19] Together, these studies lead to the current view that OCT4 uses the POUS domain to recognize its partial canonical motif in the nucleosomal DNA. However, functional studies have shown that the entire DBD region (DBDR) of OCT4 (POUS-loop-POUHD; residues 138–290, Figure 1B), termed ^{DBDR}OCT4, is essential for generating induced pluripotent stem cells.[19–21]

Here, we report the cryo-EM structures of the nucleosome containing human *LIN28B* DNA derived from the chromatin locus bound by OCT4 during cell reprogramming and its complexes with the OCT4 POUS-loop-POUHD region. Analysis of ChIP-seq and MNase data and the determination of the structure of the *ESRRB*-nucleosome bound to OCT4 suggest common structural features of nucleosome targeting by OCT4. Together with biochemical experiments, these structures suggest that multiple OCT4 molecules can target non-canonical DNA sequences in the nucleosome, which leads to the eviction of linker histone H1 and unwrapping of the nucleosomal DNA, providing insights into how OCT4 may open closed chromatin.

RESULTS

Structural determination of the *LIN28B* nucleosome bound to $^{MBP\text{-}DBDR}$OCT4

Previous biochemical studies have shown that the full-length OCT4 binds the *LIN28B* nucleosome with ~1:1 stoichiometry in the electrophoretic mobility shift assay (EMSA). [9,22,19] We tried to reconstitute the *LIN28B* nucleosome-OCT4 complex but found that the full-length OCT4 showed very low solubility and tended to aggregate when bound to the nucleosome. To improve the solubility of OCT4, we fused maltose-binding protein (MBP) to the N-terminus of the full-length OCT4 and ^{DBDR}OCT4 through a flexible linker, referred to ^{MBP}OCT4 and $^{MBP\text{-}DBDR}$OCT4, respectively. MBP has a chaperone function that helps solubilize recombinant proteins but does not bind the nucleosome.[23,24] Indeed, MBP-fused OCT4 proteins are substantially more soluble. Unexpectedly, EMSA experiments showed that multiple $^{MBP\text{-}DBDR}$OCT4, ^{DBDR}OCT4, and ^{MBP}OCT4 molecules could bind the nucleosome with the same apparent affinity, with $^{MBP\text{-}DBDR}$OCT4 displaying higher solubility (Figure S1). Also, previous studies have shown that the intrinsically disordered tail of OCT4 does not bind the nucleosome.[15] Instead, it interacts with the coactivator protein, Med1, to form a condensate in a liquid-liquid phase separation.[25] Accordingly, we chose $^{MBP\text{-}DBDR}$OCT4 in our structural study. We collected images of the nucleosome bound to scFv without and with $^{MBP\text{-}DBDR}$OCT4, using the single-chain antibody (scFv)-aided cryo-electron microscopy (cryo-EM) approach that helps prevent nucleosome dissociation.[26]

We obtained density maps of nucleosome-scFv$_2$, nucleosome-scFv$_2$-($^{MBP\text{-}DBDR}$OCT4)$_3$, and nucleosome-scFv$_2$-($^{MBP\text{-}DBDR}$OCT4)$_2$ at overall resolutions of ~2.6 Å and built the structural models (Figure 1C, Figures S2 and S3, Tables 1 and S1, Video S1).

Overall structures of the *LIN28B* nucleosome and its complexes with OCT4

In these structures, all nucleosomes are uniquely positioned with the same dyad location (Figure 1A). scFv binds on the core histone surface in the nucleosome, showing no interactions with DNA and $^{MBP\text{-}DBDR}$OCT4 (Figures S2 and S3). The structure of the free 162 bp *LIN28B* nucleosome without $^{MBP\text{-}DBDR}$OCT4 includes well-defined 149 bp DNA. In the nucleosome-scFv$_2$-($^{MBP\text{-}DBDR}$OCT4)$_3$ complex, $^{MBP\text{-}DBDR}$OCT4s bind the noncanonical DNA sequences that are in the reverse direction (relative to the canonical motifs) (Figure 1A) at three sites near the super-helical locations (SHL) −5, −1.5, and +6.5, respectively. Hereafter, we refer to the accessible sites in the nucleosomal DNA as nucleosomal sites (Nu-sites) to distinguish them from those in free DNA.

At site 1, the whole ^{DBDR}OCT4 binds the Nu-site; POUHD is wedged between the DNA and the nucleosome core, unwrapping ~25 bp nucleosomal DNA from the entry side of the nucleosome. At sites 2 and 3, only POUS recognizes the Nu-sites; the loop-POUHD region is missing, presumably flexible. In the nucleosome-scFv$_2$-($^{MBP\text{-}DBDR}$OCT4)$_2$ complex, POUS binds at sites 2 and 3 like those in the nucleosome-scFv$_2$-($^{MBP\text{-}DBDR}$OCT4)$_3$ complex. We observed weaker and broad density for ~15 bp DNA at the entry site of the nucleosome (Figure S3J), indicative of flexible conformation, likely caused by $^{MBP\text{-}DBDR}$OCT4 binding.

Specific OCT4 binding at sites 1–3

It is unexpected that at all three sites OCT4 binds the noncanonical DNA sequences rather than the canonical motifs. To confirm that OCT4 binding at sites 1–3 is sequence-specific, we used the noncanonical DNA sequences at sites 1–3 in the *LIN28B* nucleosome to substitute the nucleotides at the corresponding locations in the W601 nucleosome to obtain a mutant nucleosome, ^{W601_3sites}nucleosome. EMSA titration experiments of $^{MBP\text{-}DBDR}$OCT4 binding showed four sharp bands above the free ^{W601_3sites}nucleosome (Figure 2A). The control experiment showed that $^{MBP\text{-}DBDR}$OCT4 binding to the free W601 nucleosome leads to one sharp band above the free nucleosome (Figure 2B). Quantitative analysis of the EMSA experiments showed that $^{MBP\text{-}DBDR}$OCT4 binds ^{W601_3sites}nucleosome much stronger than the W601 nucleosome but about two times weaker than the 162 bp *LIN28B* nucleosome (Figure 2C). These results suggest that $^{MBP\text{-}DBDR}$OCT4 bind the three-sites in the *LIN28B* nucleosome specifically and a weak specific binding site for OCT4 exists in the W601 nucleosomal DNA, providing additional evidence to support that OCT4 binding to the *LIN28B* nucleosome at sites 1–3 is sequence-specific and suggests that there is a fourth binding site in the 162 bp *LIN28B* nucleosome (see later results for binding at site 0). We note that the OCT4 concentration used in our study is in the range from 0.2 to ~16 μM, which is lower than those of OCT4 and SOX2 used in the earlier structural studies.[15,16]

Interactions between ^{DBDR}OCT4 and the Nu-sites

At site 1, ^{DBDR}OCT4 binds the Nu-sites like the corresponding region of human OCT1 binding to a similar DNA sequence in a DNA fragment (Figure 3A–C and Figure S4). Notably, the fourth base pairs, different in the corresponding sequences, do not form hydrogen bonds with the DBDs. The loop region between POUS and POUHD, consisting of three Arg and two Lys residues (RK motif, residues 230–234), forms a basic patch that interacts with the DNA with a narrow minor groove consisting of an AT-track (Figure 3A and Figure S5)[20]. We found that mutation of the basic residues to Ala reduces the binding affinity and specificity between $^{MBP\text{-}DBDR}$OCT4 and the nucleosome, consistent with the earlier study that removing the positively charged residues not only disrupts their interactions with the DNA but also alters the structure of ^{DBDR}OCT4 (Figure S5B, C).[20]

At sites 2 and 3, POUS Arg186 may potentially form multiple hydrogen bonds with the bases in the Nu-sites' third and fourth base pairs (Figure 3D, E, and Figure S4). In contrast, only one hydrogen bond forms between the corresponding Arg residue in POUS and the third base of the canonical motif (ATGC) in the DNA fragment (Figure 3C). Notably, the DNA shape in the nucleosome at site 2 differs substantially from those in sites 1, 3, and the DNA fragment (Figure 3F). In addition, DNA bending by the core histones at site 2 enlarges the major groove. These results reveal that POUS Arg186, with a long side chain and the guanidino group, can form multiple hydrogen bonds with the Nu-sites with non-canonical DNA sequence. The lack of AT-tracks at sites 2 and 3 is likely the cause for the missing of the basic patch (Figure S5A). The potential formation of multiple hydrogen bonds between POUS and the Nu-sites provides additional evidence that the binding of the non-canonical motifs by ^{DBDR}OCT4 at sites 1–3 are sequence-specific.

Distinct binding modes of OCT4

In the nucleosome-scFv$_2$-($^{MBP\text{-}DBDR}$OCT4)$_3$ structure, $^{MBP\text{-}DBDR}$OCT4 binding leads to unwrapping of nucleosomal DNA and the relocation of histone H2B N-terminal tail between the two DNA gyres near site 1 (Figure 4A). The DNA conformations at sites 2 and 3 remain unchanged compared to those in the nucleosome without $^{MBP\text{-}DBDR}$OCT4 (Figure 4B). To understand why the canonical motifs, close to site 2 (Figure 1A), in the *LIN28B* nucleosome are not recognized by OCT4, we aligned the canonical motif (ATGC or AAAT) in the nucleosome structure to the corresponding motif in the DNA fragment bound to POUS or POUHD. Both POUS and POUHD clash with the core histones (Figure 4C), suggesting that the intrinsic positioning of the *LIN28B* nucleosome plays a determinant role in decoupling OCT4 recognition of the nucleosomal motifs from those in the free DNA [19]. Another unexpected finding is that the modes of OCT4 binding to the *LIN28B* nucleosome are distinct in comparison with those identified in the previous study using the W601 nucleosome as the host with the incorporation of the canonical motif as the guest (Figure 4D, E).[15] In our structures, POUS binds the major grooves at super-helical locations (SHL) −4.5, −1.5, and 6.5, and POUHD serves as a wedge to unwrap nucleosomal DNA. In contrast, in the W601-nucleosome-based structure, in which the canonical motifs of both POUS and POUHD were intentionally incorporated at the specific locations, POUS binds at either SHL −5.5 or +5.5, and POUHD is missing.

Unwrapping nucleosomal DNA by OCT4

Our structures suggest that OCT4 binding can lead to the unwrapping of the nucleosomal DNA at the entry but not the exit side of the *LIN28B* nucleosome (Figure 5A–C). The conformational change between the free nucleosome and that bound to OCT4 would lead to differences in sensitivity to micrococcal nuclease (MNase) and the distance between the entry side DNA and core histones. To verify it, we first performed micrococcal nuclease (MNase) digestion. As the ratio of $^{\text{MBP-DBDR}}$OCT4 to nucleosome increased, we found that MNase cut the DNA in the *LIN28B* nucleosome more efficiently (Figure 5D). To test the role of the POUHD domain in DNA unwrapping, we conducted a similar MNase digestion experiment using $^{\text{MBP-POUS-loop}}$OCT4 by deleting the POUHD domain. We found that deletion of the POUHD domain dramatically reduced the capability of $^{\text{MBP-DBDR}}$OCT4 to unwrap the nucleosome (Figure 5D), consistent with our structure showing that POUHD serves as a wedge in DNA unwrapping. We next conducted fluorescence resonance energy transfer (FRET) experiments with donor (Cy3) labeled at the entry or exit site of the DNA and the acceptor (Cy5) labeled at histone H2A residue 116, respectively. We measured the FRET signal by titrating the nucleosome with $^{\text{MBP-DBDR}}$OCT4. As the ratio of $^{\text{MBP-DBDR}}$OCT4 to nucleosome increased, the Cy5 FRET signal became weaker when the Cy3 donor is labeled at the entry site (Figure 5E, left). In contrast, when the Cy3 donor is labeled at the exit site, OCT4 binding did not cause change of the FRET signal (Figure 5E, right). These results indicate that OCT4 only unwraps the nucleosomal DNA at the entry site and POUHD plays an essential role, in agreement with our structures.

OCT4 binding at site 0 and eviction of H1

In the structure of the *LIN28B* nucleosome bound to three OCT4s, we identified three binding sites. Examination of the *LIN28B* DNA sequence (Figure 1A) shows that another site (termed site 0) at the entry side of the nucleosome has the identical DNA sequence as site 1, suggesting that OCT4 may also bind to site 0. This additional binding site explains the observation that there were four sharp bands above the *LIN28B*-nucleosome in the EMSA experiment (Figure S1). Structural modeling showed that OCT4 POUHD could bind the exposed motif at site 0 in the chromatosome, with POUS clashing with the linker histone globular domain and the DNA near the dyad (Figure 5F).[2] This result suggests that OCT4 might bind site 0 and inhibit H1 binding, consistent with the earlier study using an OCT4 mutant in which the five cysteine residues were replaced by other residues to avoid the formation of disulfide bonds.[22] The exact cause of the missing OCT4 at site 0 in our cryo-EM density map is unknown. We speculate that the DNA conformation at this site may be flexible since it is partially on the linker DNA, making the OCT4 unobservable.

To further verify it, we reconstituted the 187 bp *LIN28B* nucleosome without and with a site 0 mutation. We found that $^{\text{MBP-DBDR}}$OCT4 could largely replace linker histone H1 in the chromatosome at a 1:1 ratio of $^{\text{MBP-DBDR}}$OCT4 versus chromatosome in the EMSA experiments (Figure 5G). It failed to do so when site 0 was mutated to abolish $^{\text{MBP-DBDR}}$OCT4 binding at this site (Figure 5H). We solved the structure of the 187 bp nucleosome with site 0 mutation (Figures 5I and S2H–K). The high-resolution of the structure allowed us to define the dyad of the nucleosome specifically. The dyad of the nucleosome is located at the same position as in the 162 bp *LIN28B* nucleosome, indicating

the extension of the linker DNA region and site 0 mutation do not cause repositioning of the nucleosome.

We confirmed this conclusion by conducting the FRET experiments with Cy3 and Cy5 labeled on H1.4 residue 26 and H2A residue 116, respectively (Figure 5I). We found that at a 1:1 ratio of OCT4 over the chromatosome, FRET intensity decreased substantially. To further verify the role of site 0 in OCT4 binding and H1 eviction, we conducted the same FRET experiment using the chromatosome with the site 0 mutation. Indeed, we found that at a 1:1 ratio of OCT4 over the chromatosome, FRET intensity remained essentially unchanged, indicating no eviction of H1 occurred (Figure 5K).

DNA sequence features in the nucleosomes associated with OCT4

OCT4 targets many nucleosomes *in vivo* during cell reprogramming. To investigate the generality of the structural features revealed from our studies on the *LIN28B* nucleosome bound to OCT4, we analyzed the previous OCT4 ChIP-seq and MNase-seq datasets.[9,10] We created a sub-dataset consisting of sequences from −100 to +100 bp of MNase peak centers with larger than or equal to 50 bp overlap with OCT4 ChIP-seq sequences. We assume these sequences (n=63248) are the nucleosomal DNA regions bound to OCT4. We found that each of the canonical and non-canonical sequences (those observed at sites 1–3 of *LIN28B* DNA) occur with similar frequency (Figure 6A). Thus, the total non-canonical sequences occur ~two times more frequently than the canonical motifs. Furthermore, the nucleosomes comprising 5–12 non-canonical sequences dominate the population (Figure 6B). In addition, ~2% of the DNA sequences (~1265) do not contain canonical motifs, indicating that OCT4 must bind non-canonical DNA sequences in these nucleosomes. Considering both the canonical motifs and non-canonical DNA sequences, most of the DNA sequences comprise ~8–19 potential binding sites. Thus, our analysis suggests that OCT4 binding to nucleosomes at multiple sites could occur frequently in vivo.

To further explore the general features of OCT4 binding to nucleosomes with genomic DNA, we investigated the structure of OCT4 bound to the nucleosome consisting of the 168 bp *ESRRB* DNA (Figure 6C), another in vivo target of OCT4.[27] The *ESRRB* gene is a crucial regulatory element of the pluripotency network.[28] In the EMSA experiment, we observed three bands above the free nucleosome when $^{\text{MBP-DBDR}}$OCT4 was titrated into the *ESRRB*-nucleosome solution (Fig. S6A), suggesting that three $^{\text{MBP-DBDR}}$OCT4s could bind the same nucleosome specifically. We collected cryo-EM images of the $^{\text{MBP-DBDR}}$OCT4-*ESRRB* nucleosome complex following the same procedure for the $^{\text{MBP-DBDR}}$OCT4 OCT4-*LIN28B* nucleosome. We obtained two density maps for the $^{\text{MBP-DBDR}}$OCT4-*ESRRB* complex at overall resolutions of 2.8 and 3.0 Å, respectively (Fig. S6), allowing us to define the nucleosome dyad unambiguously. In one density map, two OCT4s bound the nucleosome at sites a and b, respectively (Figure 6C and Fig. S6D). At site a, POUHD bound a non-canonical DNA sequence (GGAC) at SHL −5.5. ~15 bp nucleosomal DNA nearby is unwrapped, consisting of a non-canonical DNA sequence (ATGT) for POUS, the same sequence at site 1 of the *LIN28B* nucleosome. Also, like site 1 the *LIN28B* DNA, there is one nucleotide between the GGAC and ATGT sequences. We speculate that the POUS domain may play a role in the DNA unwrapping by dynamically binding to the ATGT site

(Figure 6C). At site b, OCT4 used its POUS domain to bind the canonical DNA sequence (ATGC) at SHL 4.5 (Figure 6C). From the other density map, focused classification showed that OCT4 used the POUS domain to bind the non-canonical sequence (ATTC) at site c (Figure 6E), which has the identical sequence at site 2 in the *LIN28B* DNA.

In summary, our analysis of the MNase-seq and ChIP-seq datasets and comparison of the structures of OCT4 bound to the *LIN28B* and *ESRRB* nucleosomes suggests the following general features of nucleosome targeting by OCT4: (1) Multiple OCT4s can target the nucleosome. (2) OCT4 can bind non-canonical DNA sequences. (3) OCT4 binding can lead to the unwrapping of nucleosomal DNA, which are likely to occur in vivo.

OCT4s cooperatively open H1-condensed nucleosome array

It is hypothesized that pioneer transcription factors is capable of opening closed chromatin.[3] To test it, we chose to investigate whether OCT4 can open the H1-condensed nucleosome array containing the *LIN28B* nucleosome because of the high-quality structure of the *LIN28B* nucleosome bound to OCT4. Structural modeling shows that four partial motifs in the *LIN28B* chromatosome are accessible to OCT4 (Figure 7A). We reconstituted a H1-condensed nucleosome array consisting of 12 nucleosomes with the nucleosome repeat length of 197 bp. In this array, the *LIN28B* nucleosome is surrounded by five and six nucleosomes consisting of W601 sequence. We used the DdeI restriction enzyme to cut a specific site in the DNA in the *LIN28B* nucleosome (Figure 7 and Figure S7). We found that ^{DBDR}OCT4 enhanced the DNA digestion efficiency of the nucleosome array. Intriguingly, the digestion efficiency showed little change when the molecular ratio of ^{DBDR}OCT4 over the nucleosome array was less than two but increased sharply when the ratio is between two to four (Figure 7B, C). When the ratio was more than four, the additional OCT4 has little effect on digestion efficiency. The S-shaped curve of digestion efficiency versus the ratio of OCT4 over the nucleosome array suggests that four ^{DBDR}OCT4 molecules work cooperatively to open the chromatin target. This result agrees well with the mono-nucleosome study that four OCT4 molecules can bind to the *LIN28B* nucleosome specifically.

To further test whether OCT4 specifically binds to the *LIN28B* nucleosome in the nucleosome array and opens the nucleosome array locally or globally, we investigated the effect of OCT4 on the digestion efficiency of the nucleosome array with cutting sites at the DNA in all W601 nucleosome cores (Figure 7D), or all W601 linker DNA (Figure 7E), or the W601 nucleosome that replaced the *LIN28B* nucleosome but keep the same cutting site (Figure 7F). We found OCT4 has no effects on the digestion efficiency in these cases. Therefore, these experiments indicate that OCT4 only opens the H1-condensed nucleosome array locally at the *LIN28B* nucleosome site.

Next, we explored the role of sites 0 and 1 in opening the *LIN28B* nucleosome in the nucleosome array. We measured the mutation effects of sites 0 and 1 on the DdeI cutting efficiency of the nucleosome array (Figures 7G, H). We found that adding OCT4 did not increase the digestion efficiency of the nucleosome array with either of the mutations (Figure 7C). These results reveal OCT4 binding to both sites 0 and 1 are required for

efficient opening of the nucleosome array, consistent with the observation that OCT4s work cooperatively to open the nucleosome array.

DISCUSSION

Our finding that multiple OCT4s recognize the *LIN28B* nucleosome through noncanonical DNA sequences using both POUS and POUHD is in contrast with the current view that only POUS binds at a single Nu-site containing the canonical DNA sequence.[9,15,19] We also find that OCT4 uses two linked domains to bind and block the DNA from rewrapping, explaining the mutation results that the entire DBDR is essential for initiating cell reprogramming.[19–21] By contrast, SOX11/SOX2 unwraps nucleosomal DNA through local DNA distortion.[16] Thus, two DBDs, linked through a loop, working together to evict the linker histone and unwrap the nucleosome DNA is likely a unique feature of OCT4 for its pioneer function. It is worth noting that the previous study using the W601 nucleosome as the host shows that OCT4 loop-POUHD plays no role in either nucleosome recognition or nucleosomal DNA unwrapping.[15] Computer simulation studies suggest nucleosome dynamics may play a role in determining the binding behavior of OCT4 as the DNA in the *LIN28B* nucleosome is significantly more dynamic than that in the W601 nucleosome.[29,30] Alternatively, SOX2 binding at the adjacent site might affect the binding of OCT4 (Figure 4E).[15] Together, these results suggest that OCT4 POUS can access broad locations in the nucleosomal DNA, and the nucleosome context plays a critical role in OCT4 binding.

Our study provides insights into the hypothesis of how pioneer transcription factors may target and open closed chromatin.[3,31,32] Structures and modeling show that four accessible Nu-sites of POUS and POUHD exist in the *LIN28B* chromatosome (Figure 6A). However, chromatosomes in closed chromatin are likely to pack in a zigzagged manner (Figure S7G),[33,34] preventing direct access by OCT4. Nevertheless, transient exposure of the chromatosome could occur through intrinsic structural fluctuations,[35] allowing binding of multiple OCT4 molecules simultaneously, leading to cooperativity. For example, individual linker histones have a short residence time in condensed chromatin *in vivo*.[36] Notably, the sizes of the OCT4 DBDs and the globular domains of linker histones are comparable. As POUS can form multiple hydrogen bonds with the Nu-sites, OCT4 may use POUS to scan the condensed chromatin for initial engagement. After the initial recognition of the transiently exposed nucleosome using partial Nu-sites, OCT4 could further evict linker histones and partially unwrap the nucleosomal DNA (Figure 7). Our results suggest that this process requires OCT4s to work cooperatively.

The formation of the multi-valent binding of a pioneer transcription factor is likely to increase its overall time associated with the nucleosome through mass action,[37] facilitating the recruitment of other transcription factors for transcription activation. For multi-valent binding to occur, however, it may require higher concentration of the pioneer factor. Yamanaka's experiment on cell reprogramming involves overexpression of pioneer factors. Alternatively, higher concentration of pioneer factors could be achieved by forming condensates through liquid-liquid phase separation.[38] Recent studies have shown that OCT4 participates in liquid-liquid phase separation through the interactions of its intrinsically disordered activation domain with coactivator Med1,[25] which is required for

cell reprogramming.[39] In the *in vitro* study, the phase separation occurred at 10 μM or higher concentrations of OCT4. The local concentration of OCT4 in the condensate would be much higher. In general, most eukaryotic transcription factors bind the DNA with lower affinity because of the shorter motifs compared to those in prokaryotes.[40] Target specificity of transcription factors could be driven by their inhomogeneous 3D nuclear distribution of transcription factors and by variation in DNA binding affinity such that locally elevated transcription factor concentration allows low-affinity binding sites to be functional. Our study suggests that the number of binding sites on the nucleosomal DNA (local concentration of DNA substrate) for a transcription factor can also play a role in regulating transcription factor binding.

Finally, we speculate that the general structural features for mammalian pioneer transcription factors to engage the chromatin may include multiple Nu-sites, eviction of linker histones, and unwrapping of nucleosomal DNA. We anticipate our approach will inspire future investigations of other pioneer transcription factors bound to nucleosomes of their *in vivo* target, [31,41] providing insights into broader biological functions.

Limitations of the study

Our study provides structural insights into how OCT4 recognizes the *LIN28B* nucleosome and a working model for opening closed chromatin by OCT4 during cell reprogramming. We used the DNA binding region of OCT4 for structural study since it has a better solubility, which does not show whether the intrinsically disordered regions in the full-length protein interact with the nucleosome. Nevertheless, it is consistent with the earlier structural study that the tails are not found to interact with the nucleosome,[15] and ^{DBDR}OCT4, and the full-length OCT4 show similar apparent binding affinity (Figure S1). This study does not provide the structural information of OCT4 bound to site 0 of the nucleosomal DNA and how OCT4 evicts the linker histone. In addition, *in vivo* validation is yet to be demonstrated and may be investigated further using CRISPR/Cas9 in combination with ATAC-seq or MNase-seq.

STAR★Methods

Resource Availability

Lead Contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Yawen Bai (baiyaw@mail.nih.gov).

Materials Availability—Unique and stable reagents generated in this study are available upon request.

Data and Code Availability

- The cryo-EM reconstructions and atomic models of the *LIN28B* nucleosome and the nucleosome-$^{MBP\text{-}DBDR}$OCT4 complexes have been deposited in the Electron Microscopy Data Bank and the Protein Data Bank under the following accession codes: EMD-26261 and PDB ID 7U0J for the 162 bp *LIN28B* DNA nucleosome; EMD-27483 and PDB ID 8DK5 for the 187 bp site0 mutated *LIN28B* DNA

nucleosome; EMD-26258 and PDB ID 7U0G for the nucleosome bound three $^{MBP\text{-}DBDR}$OCT4s; EMD-26260 and PDB ID 7U0I for the nucleosome bound two $^{MBP\text{-}DBDR}$OCT4s. The cryo-EM reconstructions and atomic models of the *ESRRB* nucleosome and the nucleosome-$^{MBP\text{-}DBDR}$OCT4 complexes have been deposited in the Electron Microscopy Data Bank and the Protein Data Bank under the following accession codes: EMD-40683 and PDB ID 8SPS for the *ESRRB* nucleosome bound two $^{MBP\text{-}DBDR}$OCT4s (tight mask and mainly site b appears); EMD-40684 for the *ESRRB* nucleosome bound two $^{MBP\text{-}DBDR}$OCT4s (further classification with both site a and site b appear); EMD-40686 and PDB ID 8SPU for the *ESRRB* nucleosome bound one $^{MBP\text{-}DBDR}$OCT4 at site c; EMD-40691 for the focused classification of site c OCT4.

- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Experimental Model and Study Participant Details

Human core histone genes were gift from Dr. Hitoshi Kurumizaka, linker histone genes and OCT4 genes were synthesized. All genes were subcloned into expression plasmids and expressed in E. coli.

Method Details

Expression and purification of histones—Recombinant human histones H2A, H2A^{L116C}, H2B, H3, H4, H1.4, and H1.4 ^{K26C} were expressed individually in Escherichia coli BL21(DE3) cells as described in a previous study [2]. All mutations in our study were generated using QuikChange kit (Agilent). Briefly, E. coli cells harboring each histone expression plasmid were grown at 37 °C in 2 × YTB Broth. 0.3 mM IPTG was added to induce recombinant protein expression for 3 h at 37 °C, When OD_{600} reached around 0.6–0.8. The cells were harvested and resuspended in 50 ml of buffer A (50 mM Tris-HCl, 500 mM NaCl, 1 mM PMSF, 5% glycerol, pH 8.0), followed by sonication on ice for 60 min. The cell lysates were centrifuged at 35,000 RPM for 20 min at 4 °C. The pellet containing histones was resuspended in 50 ml of buffer A and 7 M guanidine hydrochloride. The samples were rotated for 12 h, and the supernatant was recovered by centrifugation at 35,000 RPM for 60 min at 4 °C. The supernatants were dialyzed against buffer C (5 mM Tris-HCl, pH 7.4, 2 mM 2-mercaptoethanol, 7 M urea) for three times. The supernatant was loaded to Hitrap S column chromatography (GE Healthcare). The column was washed with buffer D (20 mM sodium acetate, pH 5.2, 100 mM NaCl, 5 mM 2-mercaptoethanol, 1 mM EDTA, and 6 M urea). The histone protein was eluted by a linear gradient of 100 to 800 mM NaCl in buffer D. The purified histones were dialyzed against water for three times, and freeze-dried.

Over-expression and purification of OCT4 proteins—The plasmid (pET30a) containing the human $^{MBP\text{-}DBDR}$OCT4 or $^{MBP\text{-}POUS\text{-}loop}$OCT4 with a His-tag at the C-terminus and a Tobacco Etch Virus (TEV) recognition site (ENLYFQG/S) between MBP and OCT4 was cloned using NdeI and EcoRI restriction sites (Table S1). Ala mutants

for the basic patch residues in the loop region including R232A/R234A, R232A/K233A/R234A, R230A/K231A/R232A/K233A/R234A, which are termed $^{MBP\text{-}DBDR}OCT4^{2A}$, $^{MBP\text{-}DBDR}OCT4^{3A}$, $^{MBP\text{-}DBDR}OCT4^{5A}$, respectively. The plasmids harboring the above mutations were transformed into BL21(DE3) cells and grown in LB medium. When OD was about 0.8, 0.5 mM IPTG was added to induce protein expression at 18 °C for 18 hrs. Cells were harvested and resuspended in the buffer containing 20 mM Tris pH 7.4, 1 M NaCl, 2 mM DTT. After sonification and ultracentrifuge, the supernatant went through Nickel affinity beads. The column was washed by the buffer containing 20 mM Tris pH 7.4, 1 M NaCl, 2 mM DTT, 50 mM imidazole, then eluted by the buffer containing 20 mM Tris pH 7.4, 1 M NaCl, 2 mM DTT, 300 mM imidazole. The solution containing the fusion protein was concentrated to ~1 ml and went through gel filtration on Superdex 200 10/300 increase column (GE healthcare) equilibrated with buffer containing 20 mM Hepes pH 7.3, 500 mM NaCl, 2 mM DTT. The purified protein was concentrated to ~40 μM and stored at −80 °C. ^{DBDR}OCT4 was obtained by cutting the $^{MBP\text{-}DBDR}$OCT4 with TEV protease (Genscript) for overnight at 4 °C. The digested sample was mixed with Amylose resin (NEB) and incubated 3 hours. After centrifuge, the supernatant went through Nickel affinity beads. The beads were washed with the buffer containing 20 mM Hepes pH 7.4, 500 mM NaCl, 2 mM DTT, and the protein was eluted with the same buffer with additional 300 mM imidazole. The solution was further purified by injecting into Superdex 75 10/300 increase column (GE healthcare) equilibrated with buffer containing 20 mM Hepes pH 7.3, 500 mM NaCl, 2 mM DTT, and 5% glycerol. The purified ^{DBDR}OCT4 was concentrated and stored at −80 °C. The gene of the full-length OCT4 with an MBP-tag at N-terminus and a His-tag at C-terminus was inserted into the plasmid (pET30a). The plasmid was transformed into ArcticExpress (DE3) competent cells and grown in LB medium containing 20 μg/ml of gentamycin and 50 μg/ml of kanamycin at 30 °C. When OD was about 0.8, 0.5 mM IPTG was added to induce protein expression at 10 °C for at least 24 hrs. Protein purification followed the same procedure as $^{MBP\text{-}DBDR}$OCT4. The purified protein was concentrated to ~24 μM and stored at −80 °C.

Preparation of DNA—DNA fragments were prepared by PCR amplification, followed by ethanol precipitation and purification using the POROS column. Briefly, the PCR products were pelleted by 75% ethanol containing 0.3 M NaAc at pH 5.2. The sample was incubated for 60 min at −20 °C, followed by centrifugation. The pellet was resuspended in TE buffer. The sample was loaded to POROS column chromatography (GE Healthcare). The column was washed with buffer containing 20 mM Tris-HCl, pH 7.4, 5 mM 2-mercaptoethanol, and the DNA was eluted by a linear gradient of 0 to 2 M NaCl.

Reconstitution of nucleosomes and nucleosome arrays—Purified recombinant histones in equal stoichiometric ratio were dissolved in ~6 ml unfolding buffer (7 M guanidine-HCl, 20 mM Tris-Cl at pH 7.4, 10 mM DTT) and were dialyzed against refolding buffer (10 mM Tris-Cl at pH 7.4, 1 mM EDTA, 5 mM β-mercaptoethanol, 2 M NaCl, 0.1 mM PMSF) for about 12 hours. The mixture was centrifuged at 4000 rpm to remove any insoluble material. Soluble octamers were purified by size fractionation on a Superdex 200 gel filtration column. Purified histone octamers and DNA were mixed with a 1.3:1 ratio in high-salt buffer (2 M NaCl, 10 mM K/Na-phosphate at pH 7.4, 1 mM EDTA, 5 mM DTT).

1 ml mixture in a dialysis bag was placed in 600 mL of the high-salt buffer and dialyzed for 30 min followed by salt gradient dialysis. 3 L of a low-salt buffer (150 mM NaCl, 10 mM K/Na-phosphate at pH 7.4, 1 mM EDTA, 2 mM DTT) were gradually pumped into dialysis buffer with a flow rate of 2 ml/min for 18 h. The dialysis bag was then dialyzed against low-salt buffer for 30 min. The dialysis was done in the cold room. The sample was then incubated at 37 °C for 3–5 h. The mixture was centrifuged at 12,000 rpm to remove any insoluble material. The nucleosomes were further purified by ion-exchange chromatography (TSKgel DEAE, TOSOH Bioscience, Japan) to remove free DNA and histones. The purified nucleosomes were dialyzed against buffer containing 20 mM Hepes 7.3, 2 mM DTT.

To reconstitute 12 × 197 bp nucleosome array with a *LIN28B* nucleosome, or a *LIN28B* site 0_mut nucleosome or a *LIN28B* site 1_mut nucleosome in the middle, or a '601' nucleosome containing the Dde1 sites whose locations are the same as those in the *LIN28B* nucleosome, a pUC-18 plasmid was reconstructed with incorporation of sequentially ligated fragments of 5 × 197 bp W601 DNA, 1 × 197 bp *LIN28B* DNA or 1 × 197 bp *LIN28B* site0mut DNA, or 1 × 197 bp *LIN28B* site1mut DNA, or W601 DNA with Dde1 sites, and 6 × 197 bp W601 DNA. The resulting 12 × 197 bp DNA in the plasmid was verified by sequencing and restriction enzyme digestion. Large scale preparation and purification of the plasmid followed previous protocol [42]. Briefly, 12 × 197 bp DNA was released from the plasmid by EcoRV digestion (150 unit enzyme per 1 mg plasmid) and purified by stepwise PEG 6000 precipitation. The fractions containing the 12 × 197 bp DNA was used for nucleosome array reconstitution according to early methods.[43] Saturation of nucleosome in the reconstituted 12 × 197 bp nucleosome array was verified by SmaI digestion, which showed predominant mono-nucleosome band on a 1.2% agarose gel. Linker histone H1.4-bound nucleosome array was made by mixing 1.3-fold of H1.4 (relative to molar concentration of nucleosomes in the array) with the nucleosome array in 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1 mM DTT and 0.6 M NaCl buffer, followed by dialysis using the same buffer without NaCl.

Cryo-EM sample preparation and Data collection—Final concentration of 0.83 μM nucleosome containing human *LIN28B* DNA (162 bp), 2.5 μM scFv, and 6.6 μM $^{\text{MBP-DBDR}}$OCT4 were mixed in the buffer of 20 mM HEPES pH 7.3, 145 mM NaCl, 0.5 mM TCEP. The sample was incubated on ice for 0.5 h and then concentrated to ~2 μM (nucleosome), which was loaded onto the glow-discharged holey carbon grid (Quantifoil 300 mesh Cu R1.2/1.3). All the grids were blotted for 3 s at 14 °C and 100% relative humidity using an FEI Vitrobot Mark IV plunger before being plunge-frozen in liquid nitrogen-cooled liquid ethane. For the native 162 bp *LIN28B* nucleosome sample without OCT4, and 187 bp site 0_mut free nucleosome, equal volume (25 μL) for each 4 μM nucleosome was mixed with 12 μM scFv and the grids were prepared using the same parameter with nucleosome-$^{\text{MBP-DBDR}}$OCT4 complex. For the *ESRRB* nucleosome and OCT4 complex sample, final concentration 1.0 μM nucleosome containing human *ESRRB* DNA (168 bp), 3.0 μM scFv and 8.0 μM $^{\text{MBP-DBDR}}$OCT4 were mixed and incubated on ice for 0.5 h. Complex sample was then directly loaded onto glow-discharged holey carbon grids (Lacey 300 mesh Cu, Ted Pella). Grids were prepared using the same parameter with free nucleosome sample. Data were collected using a Titan Krios G3 electron microscope (Thermo-Fisher) operated

at 300kV. Micrographs were acquired in super-resolution mode at the nominal magnification of 81,000x with 0.528 Å pixel size using a 20-eV slit post-GIF Gatan K3 camera. The dose rate on the camera was set to 15 e^{-}/pixel/s. The total exposure time of each micrograph was 4 sec fractionated into 50 frames with 0.08 sec exposure time for each frame. Data collection were automated using the SerialEM software package [44]. A total of 7,800 micrographs were collected for the sample of the nucleosome bound to scFv and $^{MBP\text{-}DBDR}$OCT4, and a total of 2,446 micrographs were collected for the free *LIN28B* nucleosome sample. 8,235 micrographs were collected for 187 bp site 0_mut nucleosome sample. 7,027 micrographs were collected for *ESRRB* nucleosome bound to scFv and $^{MBP\text{-}DBDR}$OCT4.

Image processing—All the datasets were processed using RELION/3.1.3 and CryoSPARC v3.2 following the standard procedures (Figures S2, S3, S6).[45,46,47] The averaged images without dose weighting were used for defocus determination using CTFFIND4.1.[48] Images with dose weighting were used for particle picking and extraction. Particles were automatically picked using Gautomatch (https://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch/). Bad particles were removed by 2D classification and 3D classification in RELION using 2x binned particles. Particles were then re-extracted without binning. Several rounds of 3D classification were made using finer angular sampling rate. Selected particles were submitted to Bayesian polishing and then imported to CryoSPARC. CTF-refinement and non-uniform refinement were made to generate the final maps for model building.

Model building and structural analysis—For the free *LIN28B* nucleosome, an initial model of the nucleosome histone octamer and scFv was generated using the nucleosome structure (PDB: 7K61). They were fitted into the density map of the scFv-*LIN28B* nucleosome. The DNA was built into the map from scratch in COOT [49]. The histone octamer and scFv were optimized by manual rebuilding. The whole complex was refined using real space refinement in PHENIX.[50] For the 187 bp site 0_mut nucleosome data set, the same procedure was used except the 162 bp *LIN28B* nucleosome structure was used as the initial model. For the complexes of the *LIN28B* nucleosome bound to OCT4, the *LIN28B* nucleosome structure was used as an initial model of the nucleosome. Initial models of the POUS domain and DBDR of OCT4 were from the crystal structure of the mouse OCT4 DBDR bound to the DNA fragment (PDB: 3L1P).[51] DNA was built into the map from scratch using COOT. For the complex of the *ESRRB* nucleosome bound to OCT4, initial models of the histone octamer and scFv were generated from the *LIN28B*-nucleosome-scFv structure. DNA was built into the map from scratch. Initial model of OCT4 was generated in the same way as in the *LIN28B* nucleosome bound to OCT4 structures. Structures were optimized by manually rebuilding using COOT followed by further refinement using real space refinement in PHENIX.[52] Figures were made using UCSF Chimera[53] and PyMOL (Version 1.8, Schrödinger, LLC. DeLano Scientific).

Electrophoretic mobility shift assay—Nucleosomes at a final concentration of 0.3 μM were mixed with purified proteins. Typical binding reactions of complex formation were carried out for 30 min at room temperature in buffer containing 10 mM Hepes at pH 7.3, 60 mM NaCl, and 1 mM DTT. 10 μl of the binding reactions were analyzed on 5 % acrylamide

gels in 0.2 × TBE at 120 V for 70 minutes at 4 °C. After electrophoresis, gels were stained with ethidium bromide (EtBr) and quantified using Image J. We used Prism (Graphpad) to fit the binding data with the Hill equation. For the competition reactions between H1.4 and $^{MBP\text{-}DBDR}$OCT4 for binding of the nucleosome, the H1.4 was mixed with the nucleosome at 1.6:1 ratio, and the chaperone Nap1 at 2:1 ratio to the nucleosome was added to the mixture. The sample was incubated at room temperature for 15 min. ^{DBDR}OCT4 was added and incubated at room temperature for 20 min. The final buffer contained 10 mM Hepes at pH 7.3, 2 mM DTT, 120 mM NaCl. 10 μl of the binding reaction solution was analyzed on 5 % acrylamide gels at 120 V for 70 min in 0.2 × TBE. Gels were stained with ethidium bromide (EtBr).

Fluorescence resonance energy transfer assay—H1.4^{K26C} was labeled with Cy3 following the manufacture's protocol (Cytiva). 1 mg H1.4 ^{K26C} was dissolved in 1 mL degassed buffer containing 20 mM HEPES 7.1, 500 mM NaCl and 1 mM TCEP. One vial of Cy3 maleimide was dissolved in dimethylformamide (DMF) and mixed with H1.4 ^{K26C}. The mixture was agitated overnight at cold room. Size exclusive chromatography was used to separate the free Cy3 dye. H2A^{L116C} was labeled with Cy5 according to the previous publication [54]. 1 mg H2A^{L116C} was dissolved in 1 ml degassed buffer containing 800 mM HEPES 7.1, 1.5 mM Guanidine HCl, 1 mM TCEP. 1vial Cy5 maleimide (Cytiva, California) in DMF was mixed with H2A^{L116C} and incubated at room temperature for 5 hours. A Sephadex G-25 column was used to separate the free Cy5 dye. Fluorescent DNA fragments were produced using PCR by the Cy3 or Cy5 labeled primers (IDT). OCT4 was titrated into 200 nM-400 nM nucleosome in buffer 20 mM HEPES 7.1, 100mM NaCl, 0.5mM TCEP. In competition FRET assay, H1.4 was mixed with nucleosome with a ratio of 1.6:1 before addition of $^{MBP\text{-}DBDR}$OCT4. The chaperone Nap1 at 2:1 ratio to the nucleosome was added to prevent the non-specific binding. The fluorescence intensity was recorded using QuantaMaster (Photon Technology International). Excitation wavelength was set to 510 nm, and emission spectra were collected from 530nm to 730nm. Three independent experiments were performed.

MNase and restriction enzyme digestion assay—50 μL of 1.2 μM *LIN28B* nucleosome was mixed with increasing amounts of $^{MBP\text{-}DBDR}$OCT4 in buffer (20 mM HEPES pH 7.3, 50 mM NaCl, 2 mM DTT). Then, 10 X MNase digestion buffer and 0.3 units of MNase (NEB) was added to each reaction. Samples were incubated at 37 °C for 30 min, and the enzyme was inactivated by adding 50 mM EGTA. Samples were then incubated with proteinase K for 1 hour at room temperature. DNA purification was made by mixing 25% (v/v) phenol–chloroform–isoamyl alcohol mixture (25:24:1, sigma). After centrifugation, the top solution was harvested. The final purified samples were loaded onto an 8 % acrylamide gels at 150 V for 70 minutes at 4 °C in 0.2 x TBE. Gels were stained with ethidium bromide (EtBr). For the ^{DBDR}OCT4 facilitated digestion of nucleosome arrays assay, 150 ng of the nucleosome arrays were mixed with increasing amounts of ^{DBDR}OCT4 in digestion buffer (10 ul, 10 mM Tris pH 8.0, 60 mM NaCl, 1 mM magnesium chloride, 2 mM DTT) at room temperature. 0.1 units restriction enzyme DdeI (NEB), or 0.1 units restriction enzyme BanI (NEB), or 0.5 units restriction enzyme Sau96I (NEB) was added to each reaction solution. Samples were incubated at 37 °C for 40 min, and the enzyme was

inactivated by incubating at 65 °C for 20 min. Samples were then incubated with proteinase K at 50 °C for 60 min. DNA was purified by mixing the sample with 25 % (v/v) phenol–chloroform–isoamyl alcohol mixture (25:24:1, sigma). After centrifugation, the top solution was harvested. The final purified samples were loaded onto a 1 % agarose gel stained with SYBR Safe dye (Invitrogen). Electrophoresis was performed at 150 V in 0.2 × TBE buffer for 30 min. Three independent experiments were performed. Band intensities for the digestion product and input were measured using ImageJ to calculate digestion efficiency.[55]

Analysis of MNase-seq and ChIP-seq datasets—The previous OCT4 ChIP-seq and MNase-seq datasets[9,10] were used to create a sub-dataset for the nucleosomal DNA regions bound to OCT4. To do so, we first identified the peaks in the MNase-seq dataset and then selected sequences from −100 to +100 bp of MNase peak centers as the nucleosomal DNA regions. Next, the nucleosomal DNA sequences that have larger than or equal to 50 bp overlap with the OCT4 ChIP-seq sequence were chosen as the final DNA sequences. We then searched the non-canonical sites and the canonical motifs for POUS and POUHD in both forward and reversed directions.

Quantification and Statistical Analysis

For the apparent binding affinity measurements (Figure 2C, Figure S1E, and Figure S5C), we quantified band intensities using ImageJ 1.53, subtracted against the background and normalized to the signal of the corresponding free nucleosome without OCT4. We used Prism 9.4.1 to fit the binding data to the Hill equation: $\theta = [L]^n/\left(K_d^{app} + [L]^n\right)$. Here, θ is the fraction of nucleosome bound to OCT4. [L] is the total concentration of OCT4. K_d^{app} is the apparent binding constant. n is the Hill coefficient. For the MNase cutting assay (Figure 5D), intensity values of cutting products were quantified using ImageJ 1.53, subtracted against the background, and normalized to the total input signal. For the nucleosome array cutting assay (Figure 7C), intensity values of cutting products were quantified using ImageJ 1.53, subtracted against the background, and normalized to the signal of the cutting products of DdeI without OCT4. We performed three parallel experiments for all quantifications and used the average values for curve fitting. The error bars represent standard deviations. We conducted all statistical analysis and drew diagrams using GraphPad software Prism 9.1.0 for Mac and Microsoft Excel.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Bingquan Gao for constructing the $^{MBP\text{-}DBDR}_{OCT4}$ plasmid, Dr. Rick Huang and Ms. Allison Zeher for assistance in cryo-EM Data collection, Drs. Kenneth Zaret and Abdenour Soufi for helpful discussion, and Mr. Jacob Licht for comments on the manuscript. The cryo-EM work utilized NCI-NIH IRP Cryo-EM Consortium (NICE) microscopy resource and NIH high performance computing Biowulf system for data processing. This work is supported by the intramural research program at the Center for Cancer Research, National Cancer Institute, National Institutes of Health.

Inclusion and diversity

We support inclusive, diverse, and equitable conduct of research.

References

1. Luger K, Mader AW, Richmond RK, Sargent DF, and Richmond TJ (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389, 251–260. 10.1038/38444. [PubMed: 9305837]
2. Zhou BR, Feng H, Kale S, Fox T, Khant H, de Val N, Ghirlando R, Panchenko AR, and Bai Y (2021). Distinct Structures and Dynamics of Chromatosomes with Different Human Linker Histone Isoforms. Mol Cell 81, 166–182 e166. 10.1016/j.molcel.2020.10.038. [PubMed: 33238161]
3. Zaret KS (2020). Pioneer Transcription Factors Initiating Gene Network Changes. Annu Rev Genet 54, 367–385. 10.1146/annurev-genet-030220-015007. [PubMed: 32886547]
4. Iwafuchi-Doi M, and Zaret KS (2014). Pioneer transcription factors in cell reprogramming. Genes Dev 28, 2679–2692. 10.1101/gad.253443.114. [PubMed: 25512556]
5. Larson ED, Marsh AJ, and Harrison MM (2021). Pioneering the developmental frontier. Mol Cell 81, 1640–1650. 10.1016/j.molcel.2021.02.020. [PubMed: 33689750]
6. Balsalobre A, and Drouin J (2022). Pioneer factors as master regulators of the epigenome and cell fate. Nat Rev Mol Cell Biol 23, 449–464. 10.1038/s41580-022-00464-z. [PubMed: 35264768]
7. Takahashi K, and Yamanaka S (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676. 10.1016/j.cell.2006.07.024. [PubMed: 16904174]
8. Soufi A, Donahue G, and Zaret KS (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell 151, 994–1004. 10.1016/j.cell.2012.09.045. [PubMed: 23159369]
9. Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, and Zaret KS (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell 161, 555–568. 10.1016/j.cell.2015.03.017. [PubMed: 25892221]
10. Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, and Jones PA (2012). Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res 22, 2497–2506. 10.1101/gr.143008.112. [PubMed: 22960375]
11. Shyh-Chang N, Zhu H, Yvanka de Soysa T, Shinoda G, Seligson MT, Tsanov KM, Nguyen L, Asara JM, Cantley LC, and Daley GQ (2013). Lin28 enhances tissue repair by reprogramming cellular metabolism. Cell 155, 778–792. 10.1016/j.cell.2013.09.059. [PubMed: 24209617]
12. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science 318, 1917–1920. 10.1126/science.1151526. [PubMed: 18029452]
13. Zhu F, Farnung L, Kaasinen E, Sahu B, Yin Y, Wei B, Dodonova SO, Nitta KR, Morgunova E, Taipale M, et al. (2018). The interaction landscape between transcription factors and the nucleosome. Nature 562, 76–81. 10.1038/s41586-018-0549-5. [PubMed: 30250250]
14. Donovan BT, Chen H, Jipa C, Bai L, and Poirier MG (2019). Dissociation rate compensation mechanism for budding yeast pioneer transcription factors. Elife 8. 10.7554/eLife.43008.
15. Michael AK, Grand RS, Isbel L, Cavadini S, Kozicka Z, Kempf G, Bunker RD, Schenk AD, Graff-Meyer A, Pathare GR, et al. (2020). Mechanisms of OCT4-SOX2 motif readout on nucleosomes. Science 368, 1460–1465. 10.1126/science.abb0074. [PubMed: 32327602]
16. Dodonova SO, Zhu F, Dienemann C, Taipale J, and Cramer P (2020). Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. Nature 580, 669–672. 10.1038/s41586-020-2195-y. [PubMed: 32350470]
17. Tanaka H, Takizawa Y, Takaku M, Kato D, Kumagawa Y, Grimm SA, Wade PA, and Kurumizaka H (2020). Interaction of the pioneer transcription factor GATA3 with nucleosomes. Nat Commun 11, 4136. 10.1038/s41467-020-17959-y. [PubMed: 32811816]

18. Guan R, Lian T, Zhou BR, He E, Wu C, Singleton M, and Bai Y (2021). Structural and dynamic mechanisms of CBF3-guided centromeric nucleosome formation. Nat Commun 12, 1763. 10.1038/s41467-021-21985-9. [PubMed: 33741944]
19. Roberts GA, Ozkan B, Gachulincová I, O'Dwyer MR, Hall-Ponsele E, Saxena M, Robinson PJ, and Soufi A (2021). Dissecting OCT4 defines the role of nucleosome binding in pluripotency. Nat Cell Biol 23, 834–845. 10.1038/s41556-021-00727-5. [PubMed: 34354236]
20. Kong X, Liu J, Li L, Yue L, Zhang L, Jiang H, Xie X, and Luo C (2015). Functional interplay between the RK motif and linker segment dictates Oct4-DNA recognition. Nucleic Acids Res 43, 4381–4392. 10.1093/nar/gkv323. [PubMed: 25870414]
21. Jin W, Wang L, Zhu F, Tan W, Lin W, Chen D, Sun Q, and Xia Z (2016). Critical POU domain residues confer Oct4 uniqueness in somatic cell reprogramming. Sci Rep 6, 20818. 10.1038/srep20818. [PubMed: 26877091]
22. Echigoya K, Koyama M, Negishi L, Takizawa Y, Mizukami Y, Shimabayashi H, Kuroda A, and Kurumizaka H (2020). Nucleosome binding by the pioneer transcription factor OCT4. Sci Rep 10, 11832. 10.1038/s41598-020-68850-1. [PubMed: 32678275]
23. Chittori S, Hong J, Saunders H, Feng H, Ghirlando R, Kelly AE, Bai Y, and Subramaniam S (2018). Structural mechanisms of centromeric nucleosome recognition by the kinetochore protein CENP-N. Science 359, 339–343. 10.1126/science.aar2781. [PubMed: 29269420]
24. Kapust RB, and Waugh DS (1999). Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci 8, 1668–1674. 10.1110/ps.8.8.1668. [PubMed: 10452611]
25. Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, Li CH, Shrinivas K, Manteiga JC, Hannett NM, et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell 175, 1842–1855.e1816. 10.1016/j.cell.2018.10.042. [PubMed: 30449618]
26. Zhou BR, Yadav KNS, Borgnia M, Hong J, Cao B, Olins AL, Olins DE, Bai Y, and Zhang P (2019). Atomic resolution cryo-EM structure of a native-like CENP-A nucleosome aided by an antibody fragment. Nat Commun 10, 2301. 10.1038/s41467-019-10247-4. [PubMed: 31127102]
27. MacCarthy CM, Huertas J, Ortmeier C, Vom Bruch H, Tan DS, Reinke D, Sander A, Bergbrede T, Jauch R, Schöler HR, and Cojocaru V (2022). OCT4 interprets and enhances nucleosome flexibility. Nucleic Acids Res. 10.1093/nar/gkac755.
28. Festuccia N, Dubois A, Vandormael-Pournin S, Gallego Tejeda E, Mouren A, Bessonnard S, Mueller F, Proux C, Cohen-Tannoudji M, and Navarro P (2016). Mitotic binding of Esrrb marks key regulatory regions of the pluripotency network. Nat Cell Biol 18, 1139–1148. 10.1038/ncb3418. [PubMed: 27723719]
29. Huertas J, MacCarthy CM, Scholer HR, and Cojocaru V (2020). Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding. Biophys J 118, 2280–2296. 10.1016/j.bpj.2019.12.038. [PubMed: 32027821]
30. Tan C, and Takada S (2020). Nucleosome allostery in pioneer transcription factor binding. Proc Natl Acad Sci U S A 117, 20586–20596. 10.1073/pnas.2005500117. [PubMed: 32778600]
31. Cirillo LA, McPherson CE, Bossard P, Stevens K, Cherian S, Shim EY, Clark KL, Burley SK, and Zaret KS (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. EMBO J 17, 244–254. 10.1093/emboj/17.1.244. [PubMed: 9427758]
32. Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, and Zaret KS (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol Cell 9, 279–289. 10.1016/s1097-2765(02)00459-8. [PubMed: 11864602]
33. Garcia-Saez I, Menoni H, Boopathi R, Shukla MS, Soueidan L, Noirclerc-Savoye M, Le Roy A, Skoufias DA, Bednar J, Hamiche A, et al. (2018). Structure of an H1-Bound 6-Nucleosome Array Reveals an Untwisted Two-Start Chromatin Fiber Conformation. Mol Cell 72, 902–915 e907. 10.1016/j.molcel.2018.09.027. [PubMed: 30392928]
34. Song F, Chen P, Sun D, Wang M, Dong L, Liang D, Xu RM, Zhu P, and Li G (2014). Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. Science 344, 376–380. 10.1126/science.1251413. [PubMed: 24763583]

35. Zhou BR, Jiang J, Ghirlando R, Norouzi D, Sathish Yadav KN, Feng H, Wang R, Zhang P, Zhurkin V, and Bai Y (2018). Revisit of Reconstituted 30-nm Nucleosome Arrays Reveals an Ensemble of Dynamic Structures. Journal of molecular biology 430, 3093–3110. 10.1016/j.jmb.2018.06.020. [PubMed: 29959925]
36. Misteli T, Gunjan A, Hock R, Bustin M, and Brown DT (2000). Dynamic binding of histone H1 to chromatin in living cells. Nature 408, 877–881. 10.1038/35048610. [PubMed: 11130729]
37. Tang X, Li T, Liu S, Wisniewski J, Zheng Q, Rong Y, Lavis LD, and Wu C (2022). Kinetic principles underlying pioneer function of GAGA transcription factor in live cells. Nat Struct Mol Biol 29, 665–676. 10.1038/s41594-022-00800-z. [PubMed: 35835866]
38. Alberti S, Gladfelter A, and Mittag T (2019). Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. Cell 176, 419–434. 10.1016/j.cell.2018.12.035. [PubMed: 30682370]
39. Wang J, Yu H, Ma Q, Zeng P, Wu D, Hou Y, Liu X, Jia L, Sun J, Chen Y, et al. (2021). Phase separation of OCT4 controls TAD reorganization to promote cell fate transitions. Cell Stem Cell 28, 1868–1883.e1811. 10.1016/j.stem.2021.04.023. [PubMed: 34038708]
40. Kribelbauer JF, Rastogi C, Bussemaker HJ, and Mann RS (2019). Low-Affinity Binding Sites and the Transcription Factor Specificity Paradox in Eukaryotes. Annu Rev Cell Dev Biol 35, 357–379. 10.1146/annurev-cellbio-100617-062719. [PubMed: 31283382]
41. Fernandez Garcia M, Moore CD, Schulz KN, Alberto O, Donague G, Harrison MM, Zhu H, and Zaret KS (2019). Structural Features of Transcription Factors Associating with Nucleosome Binding. Mol Cell 75, 921–932.e926. 10.1016/j.molcel.2019.06.009. [PubMed: 31303471]
42. Zhou BR, and Bai Y (2021). Preparation of scFv stabilized chromatosomes for single-particle cryo-EM structure determination. STAR Protoc 2, 100396. 10.1016/j.xpro.2021.100396. [PubMed: 33786462]
43. Dorigo B, Schalch T, Kulangara A, Duda S, Schroeder RR, and Richmond TJ (2004). Nucleosome arrays reveal the two-start organization of the chromatin fiber. Science 306, 1571–1573. 10.1126/science.1103124. [PubMed: 15567867]
44. Mastronarde DN (2005). Automated electron microscope tomography using robust prediction of specimen movements. J Struct Biol 152, 36–51. 10.1016/j.jsb.2005.07.007. [PubMed: 16182563]
45. Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJ, Lindahl E, and Scheres SH (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife 7. 10.7554/eLife.42166.
46. Punjani A, Rubinstein JL, Fleet DJ, and Brubaker MA (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods 14, 290–296. 10.1038/nmeth.4169. [PubMed: 28165473]
47. Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, and Agard DA (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods 14, 331–332. 10.1038/nmeth.4193. [PubMed: 28250466]
48. Rohou A, and Grigorieff N (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol 192, 216–221. 10.1016/j.jsb.2015.08.008. [PubMed: 26278980]
49. Emsley P, Lohkamp B, Scott WG, and Cowtan K (2010). Features and development of Coot. Acta crystallographica. Section D, Biological crystallography 66, 486–501. 10.1107/S0907444910007493. [PubMed: 20383002]
50. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica. Section D, Biological crystallography 66, 213–221. 10.1107/S0907444909052925. [PubMed: 20124702]
51. Esch D, Vahokoski J, Groves MR, Pogenberg V, Cojocaru V, Vom Bruch H, Han D, Drexler HC, Arauzo-Bravo MJ, Ng CK, et al. (2013). A unique Oct4 interface is crucial for reprogramming to pluripotency. Nat Cell Biol 15, 295–301. 10.1038/ncb2680. [PubMed: 23376973]
52. Liebschner D, Afonine PV, Baker ML, Bunkoczi G, Chen VB, Croll TI, Hintze B, Hung LW, Jain S, McCoy AJ, et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D Struct Biol 75, 861–877. 10.1107/S2059798319011471. [PubMed: 31588918]

53. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25, 1605–1612. 10.1002/jcc.20084. [PubMed: 15264254]

54. Shimko JC, North JA, Bruns AN, Poirier MG, and Ottesen JJ (2011). Preparation of fully synthetic histone H3 reveals that acetyl-lysine 56 facilitates protein binding within nucleosomes. Journal of molecular biology 408, 187–204. 10.1016/j.jmb.2011.01.003. [PubMed: 21310161]

55. Schneider CA, Rasband WS, and Eliceiri KW (2012). NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9, 671–675. 10.1038/nmeth.2089. [PubMed: 22930834]

Highlights

- High-resolution structures of the *LIN28B* nucleosome bound to three OCT4s
- All OCT4s recognize non-canonical DNA sequences in the nucleosome
- One OCT4 POUHD serves as a wedge to unwrap ~25 bp nucleosomal DNA
- OCT4s work cooperatively to open H1-condensed nucleosome array

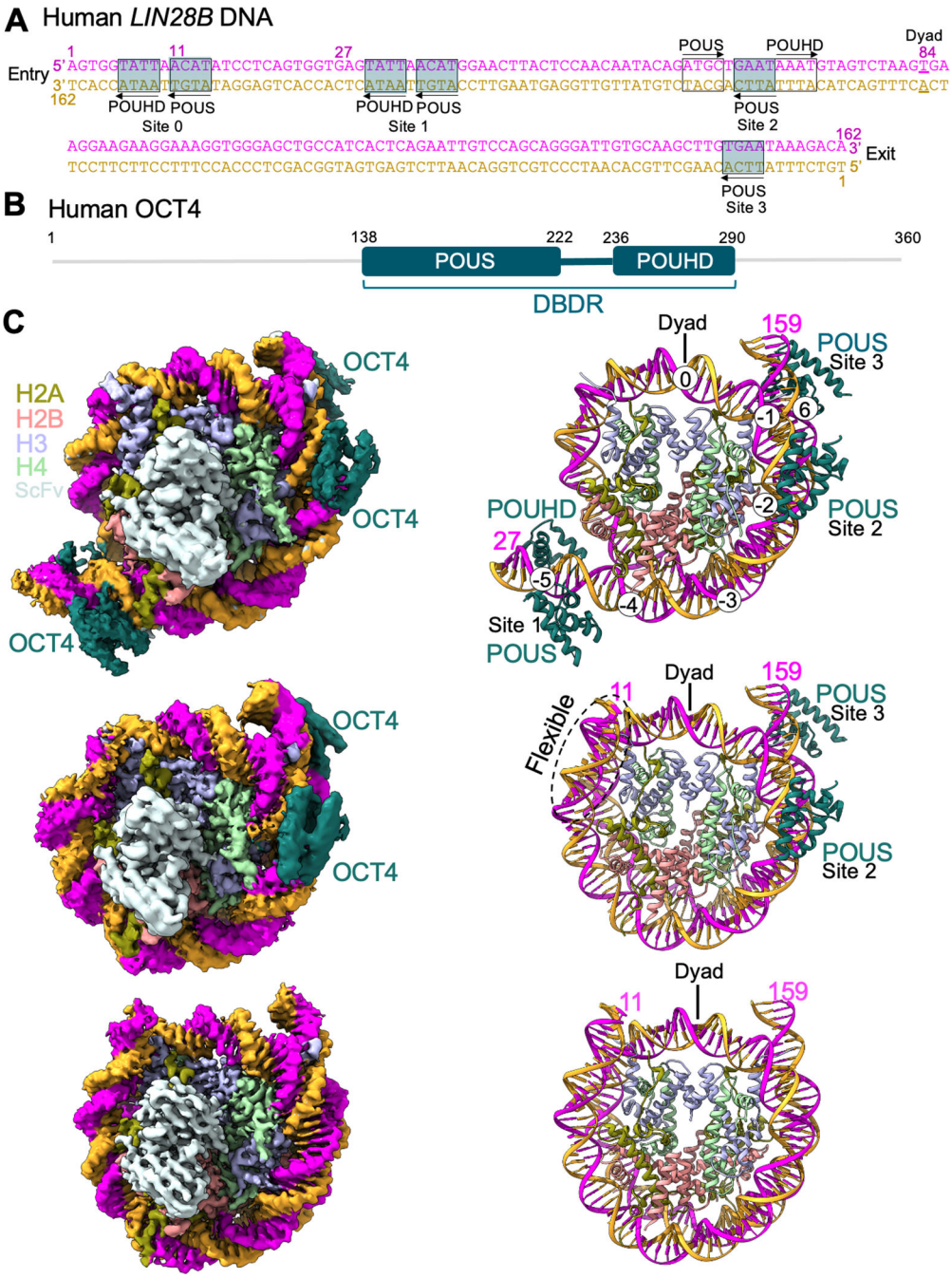


Figure 1. Overall structures of the *LIN28B* nucleosome bound to $^{\text{MBP-DBDR}}$OCT4

(A) The *LIN28B* DNA sequence. The boxes with transparent shade in cyan show the noncanonical DNA motifs for POUS and POUHD. The open boxes show the canonical motifs for POUS and POUHD, respectively. Arrows indicate the direction of the motifs. The base pair at the dyad is highlighted with the underline.

(B) Diagram illustration of the domain organization of human OCT4 protein with the residue numbers that mark the boundaries of the domains.

(C) Density maps (left) of the *LIN28B* nucleosome bound to three (top), two (middle) and 0 $^{\text{MBP-DBDR}}$OCT4s and their corresponding structural models (right). The two DNA strands are in magenta and orange, respectively. The scFv molecules have been omitted in the structural models for clarity.

See also Figures S1–3 and Table 1 and Video S1.

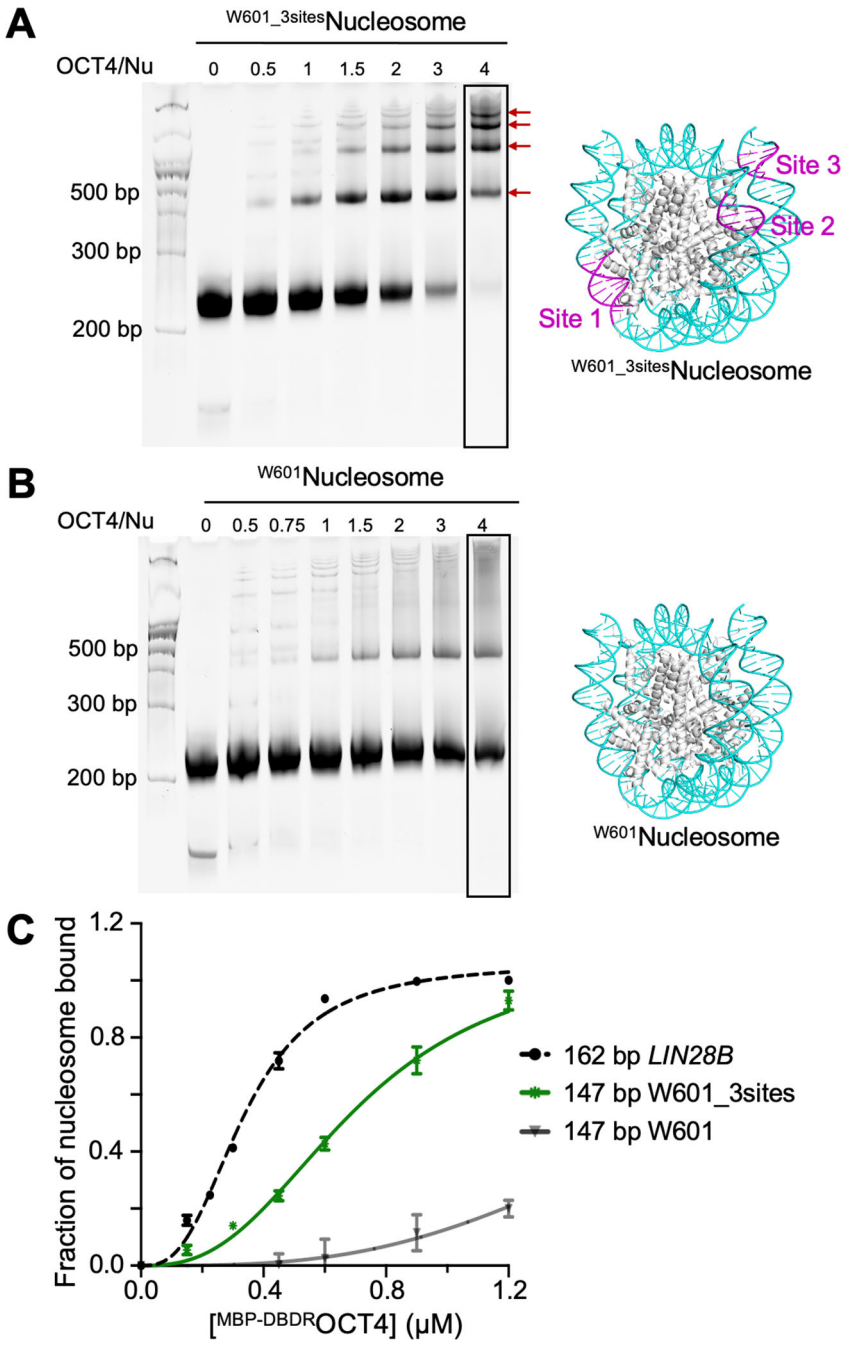


Figure 2. OCT4 binding at sites 1–3 are sequence specific

(A) EMSA showing the binding of $^{MBP\text{-}DBDR}$OCT4 (OCT4) to the $^{W601\text{-}3sites}$nucleosome (Nu) (left). The red arrows indicate the sharp bands. The three sites in the W601 nucleosome structure are shown in magenta (right).

(B) EMSA showing the binding of $^{MBP\text{-}DBDR}$OCT4 (OCT4) to the ^{W601}nucleosome (Nu) (left) with the W601 nucleosome structure (right).

(C) Plots showing the bindings of $^{MBP\text{-}DBDR}$OCT4 to the nucleosomes consisting of *LIN28B* W601_3sites, and W601 DNA. The error bars represent standard deviations from three independent experiments. Apparent K_d values (K_d^{app}) and Hill coefficients (n) were

obtained by fitting the data to Hill equation: 0.34 ± 0.04 μM (n=2.8) and 0.71 ± 0.04 μM (n=2.7) for the *LIN28B* nucleosome and the $^{W601\text{-}3sites}$nucleosome, respectively. The data for the W601 nucleosome could not be fit reliably due to very weak binding.
See also Figure S1.

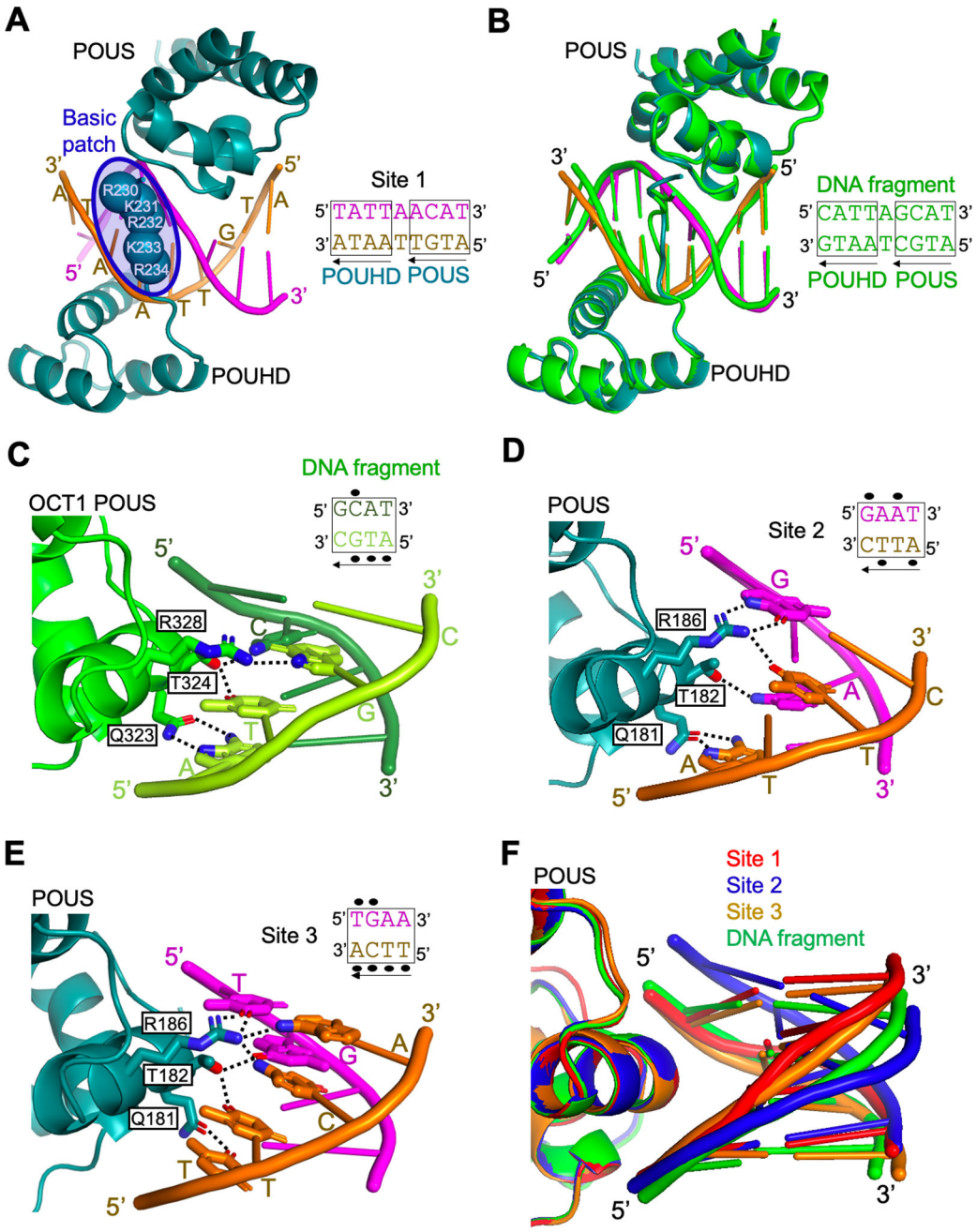


Figure 3. Interactions between OCT4 and the noncanonical motifs

(A) Structure of $^{MBP\text{-}DBDR}$OCT4 bound to Nu-motif at site 1. The basic patch residues that interact with the minor groove of DNA are represented with spheres centered at the backbone Cα carbon atoms.

(B) Alignment of the structures of $^{MBP\text{-}DBDR}$OCT4 bound to site 1 in the nucleosome and human OCT1 DBDR bound to the DNA fragment in the crystal structure (PDB ID: 1GT0).

(C) Illustration of hydrogen bonds (dashed lines) formed between POUS and the canonical DNA motif (ATGC) in the DNA fragment (PDB ID: 1GT0). They were drawn based on the distance (< 3.2 Å) between a pair of polar atoms. The side chains are shown in sticks.

The nitrogen and oxygen atoms involved in the hydrogen bond formation are colored in blue and red, respectively. The black ovals close to the bases of the sequences indicate their involvement in hydrogen bond formation.

(D) Illustration of the potential formation of hydrogen bonds (dashed lines) between POUS and sites 2 DNA. They may not form simultaneously due to angle restriction.

(E) Illustration of the potential formation of hydrogen bonds (dashed lines) between POUS and site 3.

(F) Comparison of the conformations of POUS bound Nu-sites at sites 1–3 and the canonical motif in the DNA fragment. The structures are aligned on the POUS domains.

See also Figures S4 and S5.

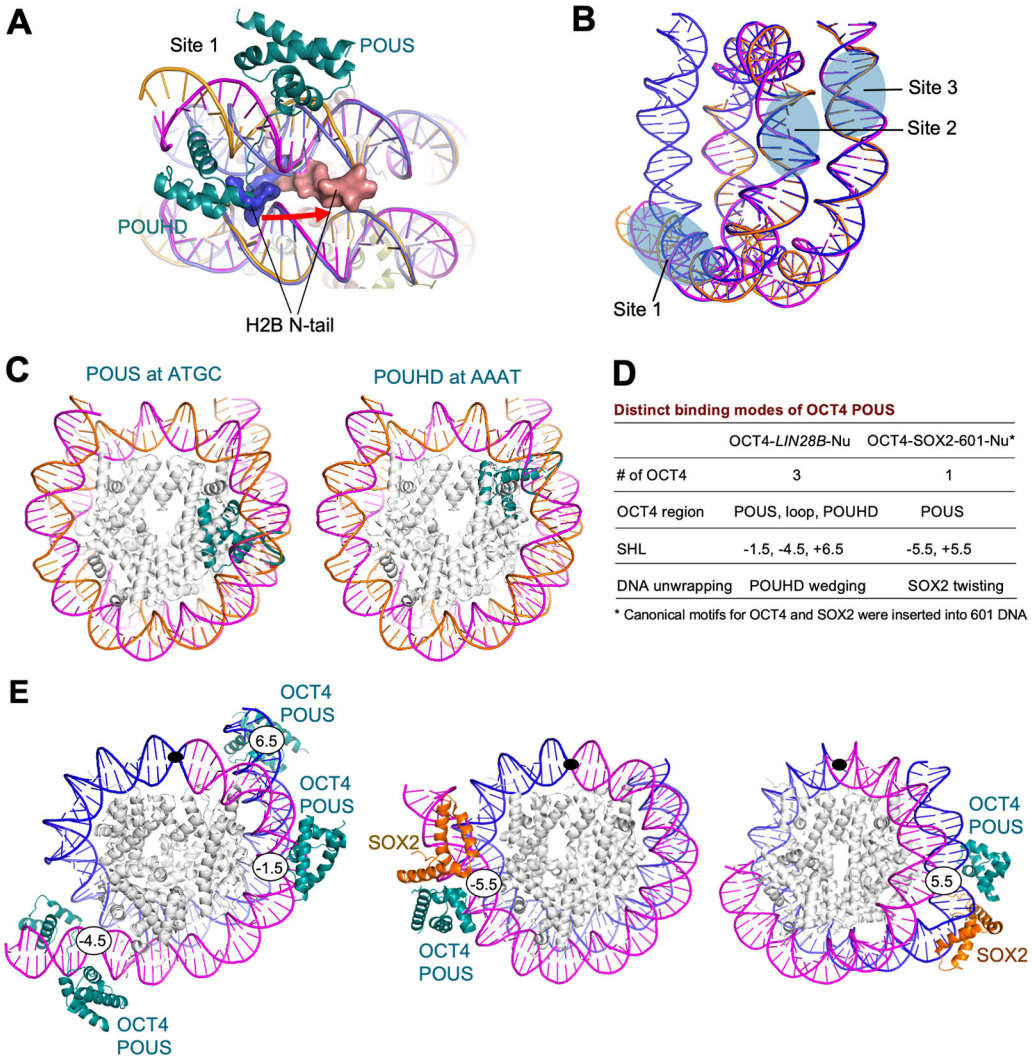


Figure 4. Conformational changes and distinct binding modes of OCT4

(A) The N-terminal tail of H2B in the free nucleosome (blue) near site 1 relocates to a new position (orange) when site 1 is bound by OCT4.

(B) Overlay of the nucleosomal DNA in the nucleosome-scFv$_2$-($^{\text{MBP-DBDR}}$OCT4)$_3$ complex (orange) and the nucleosome (blue) with the highlighted sites 1–3.

(C) Modeling of OCT4 POUS and POUHD binding to their canonical motifs (ATGC and AAAT) in the *LIN28B* nucleosome. Each motif in the nucleosome is aligned with the corresponding motif in the crystal structure of the DNA fragment bound to either POUS or POUHD.

(D) Summary of the differences in OCT4 binding modes between our structure and the two structures in the W601-based nucleosome (see E).

(E) Comparison of our structure (left) and the two structures of W601-based nucleosomes (PDB IDs: 6T90 and 6YOV) (middle and right). The filled ovals indicate the nucleosome dyad. The numbers indicate the SHLs.

See also Figure S3.

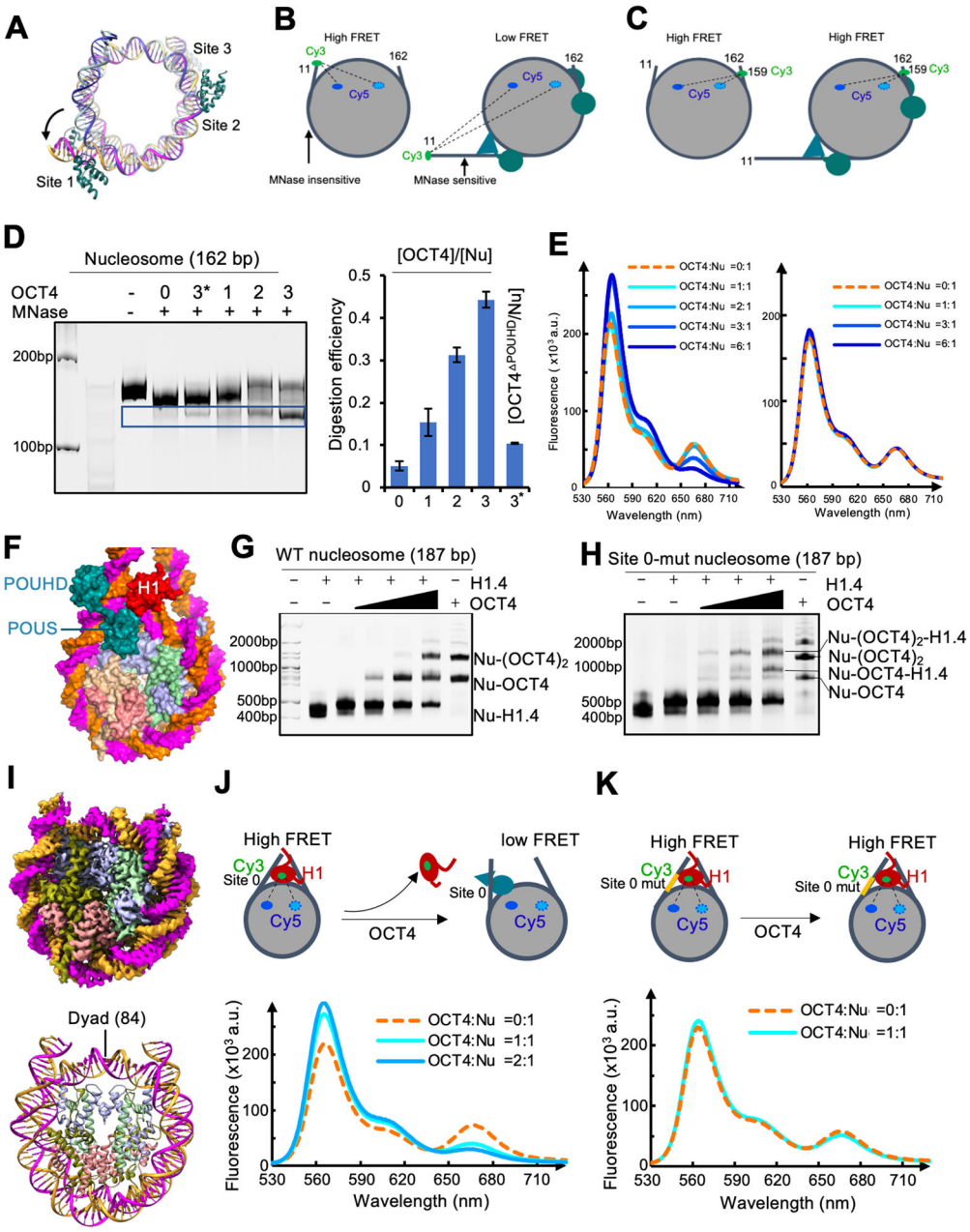


Figure 5. Biochemical studies of DNA unwrapping, site 0 binding, and H1 eviction

(A) Comparison of the DNA structures in the *LIN28B* nucleosome alone (blue) and that bound to three $^{MBP\text{-}DBDR}$OCT4s (magenta and orange), showing DNA unwrapping by OCT4.

(B) Anticipated sensitivity to MNase digestion and FRET signal change with Cy3 labeled at the entry site and Cy5 labeled on histone H2A residue 116 after OCT4 binding to the nucleosome.

(C) No FRET signal change is expected to occur with Cy3 labeled at the exit site and Cy5 labeled on histone H2A residue 116 after OCT4 binding to the nucleosome.

(D) Results of MNase digestion of the 162 bp *LIN28B* nucleosome in the presence of $^{\text{MBP-DBDR}}$OCT4 or $^{\text{MBP-POUS-loop}}$OCT4 at various concentrations (Left). The numbers indicate the ratio of $^{\text{MBP-DBDR}}$OCT4 over the nucleosome. The number with asterisk is for $^{\text{MBP-POUS-loop}}$OCT4 (POUHD deletion). The blue rectangular highlights the DNA fragments cut off from the nucleosome by MNase. Quantification of their intensity over the input is shown in the bar graph (Right). Error bars represent standard deviations from three independent experiments.

(E) Results of FRET experiments with Cy3 labeled at the entry (left) and exit (right) sites of the DNA and Cy5 labeled at histone H2A residue 116 at different molar ratios of OCT4 over the nucleosome. FRET signal (Cy5 emission) is around 670 nm.

(F) Modeling of binding of $^{\text{DBDR}}$OCT4 to site 0 in the chromatosome by aligning the *LIN28B* nucleosome with the chromatosome structure (PDB ID: 7K5Y) using the core histones and by aligning the DBDR bound to the site 1 DNA to site 0.

(G) Results of EMSA experiments for $^{\text{MBP-DBDR}}$OCT4 binding to the chromatosome. The concentrations of the nucleosome and H1.4 are 0.3 μM and 0.48 μM, respectively. [$^{\text{MBP-DBDR}}$OCT4]/[nucleosome] is 0.25, 0.5, 1.0, or 2.0.

(H) EMSA showing that binding of $^{\text{MBP-DBDR}}$OCT4 to the chromatosome containing site 0 mutation. The concentrations of the nucleosome and H1.4 are 0.3 and 0.48 μM, respectively. [$^{\text{MBP-DBDR}}$OCT4]/[nucleosome] is 0.25, 0.5, 1.0, or 2.0. The identities of the shifted bands have been surmised based on their mobilities.

(I) Structure of the *LIN28B* nucleosome with 187 bp DNA and site 0 mutation, showing the same dyad as the 162 bp *LIN28B* nucleosome. Top: density map. Bottom: ribbon representation of the nucleosome structure.

(J) FRET signals of $^{\text{MBP-DBDR}}$OCT4 binding to the chromatosome containing Cy3 labeled on H1.4 at residue 26 and Cy5 at histone H2A residue 116. FRET signal (Cy5 emission) is around 670 nm.

(K) Same as (H) except the DNA consists of site 0 mutation.

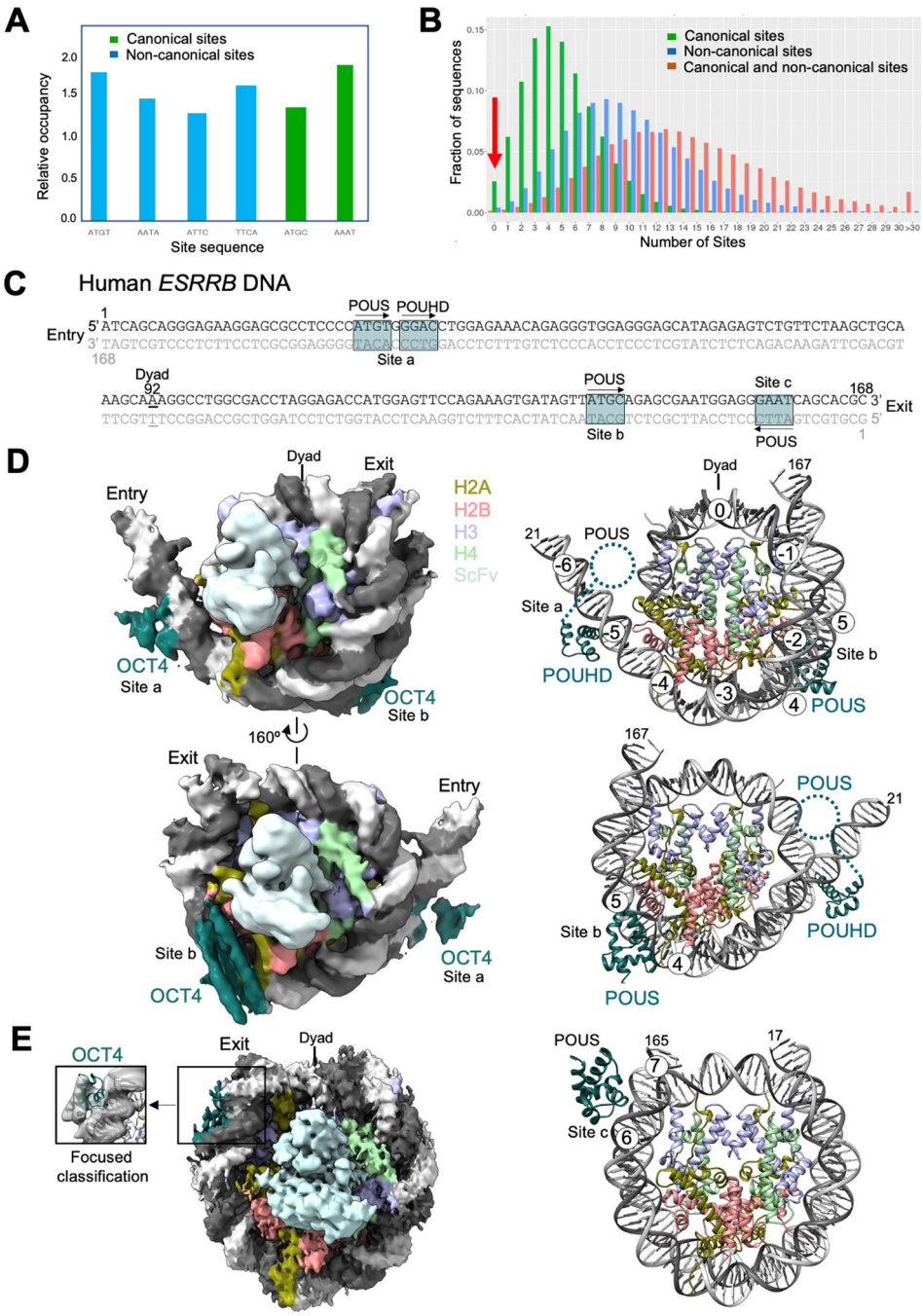


Figure 6. Features of other nucleosomes associated with OCT4

(A) Average number of the canonical and non-canonical DNA sites in the sub-dataset (n=63248).

(B) Distributions of the fraction of the nucleosomal DNA sequences containing N sites versus the number, N, of canonical and/or non-canonical sites in the sub-dataset. The red arrow highlights the fraction of sequences that do not contain a canonical motif.

(C) Human 168 bp *ESRRB* DNA sequence. The boxes indicate the DNA sequences that are bound to OCT4 in the cryo-EM structures (see below) and the arrows indicate the direction of the DNA sequences.

(D) The density map for the *ESRRB*-nucleosome bound to two OCT4 molecules at sites a and b (left) and the corresponding structural model (right). The numbers in the circles indicate super-helical locations. The scFv molecules have been omitted in the structural models for clarity (also in panel E).

(E) The density map for the *ESRRB*-nucleosome bound to one OCT4 molecule at site c (left) and the corresponding structural model (right).

See also Figure S6.

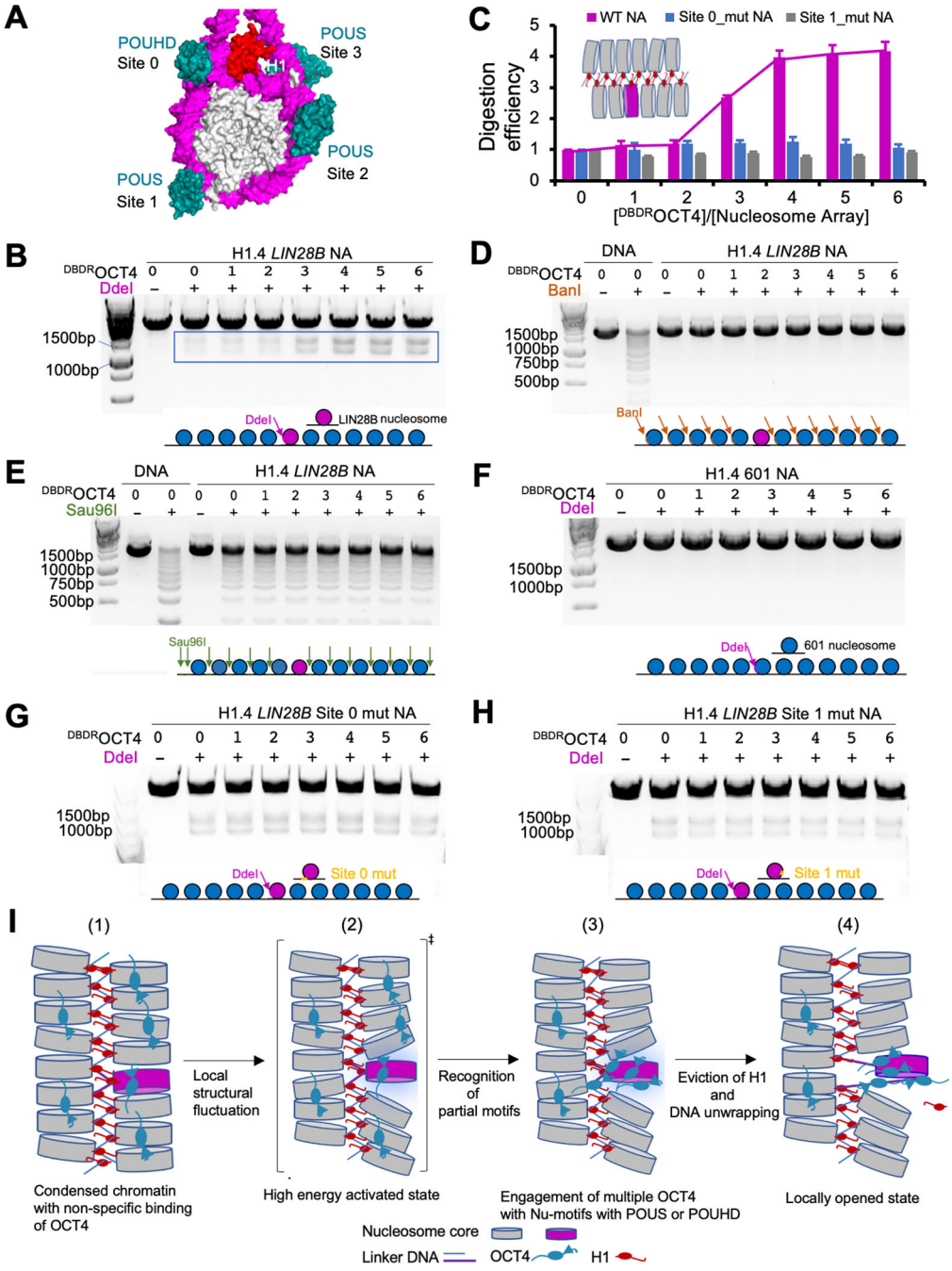


Figure 7. OCT4 opens H1-condensed nucleosome array containing the *LIN28B* nucleosome

(A) Illustration of four sites with partial motifs accessible to either POUS or POUHD in the *LIN28B* chromatosome.

(B) Digestion of the nucleosome array by DdeI enzyme with increasing ratios of ^{DBDR}OCT4 over the nucleosome array.

(C) The relative digestion efficiency of the linker histone H1-condensed nucleosome arrays consisting of wild type *LIN28B* DNA and those with sites 0 and 1 mutations (see G

and H later), respectively. Error bars represent standard deviations from three independent experiments.

(D) Digestion of the nucleosome array with BanI that cut the sites in the W601 nucleosome core.

(E) Digestion of the nucleosome array with Sau96I that cut the sites in the linker DNA.

(F) Digestion of the nucleosome array with the DdeI enzyme that cut the site within the W601 nucleosome that replaced the *LIN28B* nucleosome.

(G) Same as (B) except the DNA consists of site 0 mutation.

(H) Same as (B) except the DNA consists of site 1 mutation.

(I) Diagram illustration of a working model explaining the local opening of H1-condensed nucleosome array by OCT4. (1) Non-specific binding of OCT4 in searching for the *LIN28B* chromatosome. (2) Intrinsic stochastic fluctuations of the condensed chromatin structure transiently expose the *LIN28B* chromatosome. The parathesis with the ‡ indicate an activated high-energy state (analogues to the transition state in a chemical reaction or a high-energy intermediate state). (3) Recognition of the Nu-sites by POUS and POUHD. (4) Eviction of linker histone H1, unwrapping of the nucleosomal DNA, and stabilization of the locally opened chromatin by OCT4.

See also Figure S7.

Table 1.

Data collection and model refinement parameters. Related to Figure 1, 5, 6.

Data collection and processing				
Magnification	81,000			
Voltage (kV)	300			
Exposure time (s/frame)	0.08			
Number of frames	50			
Electron exposure (e-/Å^2)	53.8			
Defocus range (μm)	−1.0 ~ −2.0			
Pixel size (Å)	0.528			
Symmetry imposed	C1			
	***LIN28B* Nuc-ScFv$_2$ PDB ID: 7U0J EMDB ID: 26261**	**187bp site0_mut *LIN28B* Nuc-ScFv$_2$ PDB ID: 8DK5 EMDB ID: 27483**	**Nuc-scFv$_2$-($^{MBP\text{-}DBDR}$OCT4)$_2$ PDB ID: 7U0I EMDB ID: 26260**	**Nuc-scFv$_2$-($^{MBP\text{-}DBDR}$OCT4)$_3$ PDB ID: 7U0G EMDB ID: 26258**
Initial particle images (no.)	1,466,167	1,813,056	1,069,366	1,069,366
Final particle images (no.)	359,203	49,406	88,819	112,532
Map resolution (Å)	2.7	2.7	2.6	2.6
FSC threshold	0.143	0.143	0.143	0.143
Map resolution range	2.3–8.4	2.3–11.1	2.4–7.1	2.4–9.1
Map sharpening *B* factor (Å^2)	−86	−80	−20	−20
Refinement				
Non-hydrogen atoms	15,622	15,621	16,888	17,327
Protein residues	1,210	1,210	1,368	1,510
Nucleotide	298	298	298	266
***B* factors (Å^2)**				
Protein	62.25	61.98	79.43	97.67
Nucleic acids	98.79	82.06	120.09	110.25
R.m.s deviations				
Bond lengths (Å)	0.008	0.011	0.010	0.006
Bond angles (°)	0.930	1.114	0.862	0.777
Validation				
MolProbity score	1.66	1.83	2.09	1.94
Clashscore	6.00	9.46	14.30	14.52
Poor rotamers (%)	0.1	0.1	0.2	0.1
Ramachandran Plot				
Favored (%)	95.28	95.24	95.98	96.14
Allowed (%)	4.64	4.64	3.88	3.79
Disallowed (%)	0.08	0.08	0.14	0.07
Magnification	81,000			
Voltage (kV)	300			
Exposure time (s/frame)	0.08			

Data collection and processing				
Number of frames	50			
Electron exposure (e-/Å^2)	53.8			
Defocus range (μm)	−1.0 ~ −2.0			
Pixel size (Å)	0.528			
Symmetry imposed	C1			
	ESRRB **Nuc-ScFv$_2$-($^{MBP\text{-}DBDR}$OCT4)$_2$-site b PDB ID: 8SPS EMDB ID: 40683**	***ESRRB*** **Nuc-ScFv$_2$-($^{MBP\text{-}DBDR}$OCT4)$_2$-site a+b EMDB ID: 40684**	***ESRRB*** **Nuc-ScFv$_2$-($^{MBP\text{-}DBDR}$OCT4(site c) PDB ID: 8SPU EMDB ID: 40686**	**Site c $^{MBP\text{-}DBDR}$OCT4 (focused refinement) EMDB ID: 40691**
Initial particle images (no.)	5,472,720	5,472,720	5,472,720	5,472,720
Final particle images (no.)	68,790	38,160	110,991	30,272
Map resolution (Å)	3.0	4.2	2.8	7.9
FSC threshold	0.143	0.143	0.143	0.143
Map resolution range	2.3–7.0	3.5–11.0	2.3–8.0	
Map sharpening *B* factor (Å^2)	−40	−20	−40	
Refinement				
Non-hydrogen atoms	16,141		16,181	
Protein residues	1,285		1,285	
Nucleotide	294		298	
***B* factors (Å^2)**				
Protein	78.47		74.02	
Nucleic acids	236		150	
R.m.s deviations				
Bond lengths (Å)	0.008		0.008	
Bond angles (°)	0.888		0.905	
Validation				
MolProbity score	1.86		1.85	
Clashscore	10.52		10.56	
Poor rotamers (%)	0.5		0.47	
Ramachandran Plot				
Favored (%)	95.39		95.55	
Allowed (%)	4.53		4.37	
Disallowed (%)	0.08		0.08	

Key resource table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Escherichia coli BL21-CodonPlus (DE3)-RIPL Competent Cells	Agilent	Cat#230280
ArcticExpress (DE3) Competent Cells	Agilent	Cat#230192
Chemicals, peptides, and recombinant proteins		
HiTrap SP HP cation exchange chromatography column	GE Healthcare	Cat#17115101
Ni-NTA Agarose	QIAGEN	Cat#30230
TSKgel DEAE-5PW column	TOSOH	Cat#0007574
Superdex 200 increase 10/300 GL column	GE Healthcare	Cat#28990944
Superdex 75 increase 10/300 GL column	GE Healthcare	Cat#29148721
Amylose resin	New England Biolabs	Cat#E8021L
POROS GOPURE 50HQ colume	ThermoFisher	Cat#4448878
Micrococcal Nuclease	New England Biolabs	Cat#M0247S
DdeI enzyme	New England Biolabs	Cat#R0175S
Deposited data		
162 bp LIN28B nucleosome	This paper	EMDB: EMD-26261, PDB ID: 7U0J
187 bp site 0 mutated LIN28B nucleosome	This paper	EMDB: EMD-27483, PDB ID: 8DK5
LIN28B nucleosome bound two $^{MBP\text{-}DBDR}$OCT4s	This paper	EMDB: EMD-26260, PDB ID: 7U0I
LIN28B nucleosome bound three $^{MBP\text{-}DBDR}$OCT4s	This paper	EMDB: EMD-26258, PDB ID: 7U0G
ESRRB nucleosome bound two $^{MBP\text{-}DBDR}$OCT4s (tight mask with mainly site b appears)	This paper	EMDB: EMD-40683, PDB ID: 8SPS
ESRRB nucleosome bound two $^{MBP\text{-}DBDR}$OCT4s (further classification with both site a and site b appear)	This paper	EMDB: EMD-40684
ESRRB nucleosome bound $^{MBP\text{-}DBDR}$OCT4 at site c	This paper	EMDB: EMD-40686, PDB ID: 8SPU
Focused classification of site c OCT4	This paper	EMDB: EMD-40691
CryoEM structure of nucleosome bound ScFv	Zhou et al., 2019	PDB ID: 7K61
Crystal structure of mouse OCT4 bound free DNA	Esch et al., 2013	PDB ID: 3L1P
Chip-seq data of OCT4	Soufi et al., 2012	GEO: GSE36570
MNase-seq data	Kelly et al., 2012	GEO: GSM543311
Original gel images	This paper	doi:10.17632/fg8xx3rh8k.1
Oligonucleotides		
Oligonucleotides used in this study	This paper	Table S1
Recombinant DNA		
pET42b-H2A	This paper	N/A
pET42b-H2A^{L116C}	This paper	N/A
pET42b-H2B	This paper	N/A
pET21b-H3	This paper	N/A
pET21b-H4	This paper	N/A
pET42b-H1.4-his6	This paper	N/A

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pET42b-H1.4^{K26C}-his6	This paper	N/A
pET30a- $^{MBP\text{-}DBDR}$OCT4	This paper	N/A
pET30a- $^{MBP\text{-}POUS\text{-}loop}$OCT4	This paper	N/A
pET30a- $^{MBP\text{-}DBDR}$OCT4^{2A}	This paper	N/A
pET30a- $^{MBP\text{-}DBDR}$OCT4^{3A}	This paper	N/A
pET30a- $^{MBP\text{-}DBDR}$OCT4^{5A}	This paper	N/A
pET30a- ^{MBP}OCT4	This paper	N/A
pUC18–12x nucleosome array	This paper	N/A
pUC18–12x nucleosome array site0mut	This paper	N/A
pUC18–12x nucleosome array site1mut	This paper	N/A
pET15b-scFv	Zhou et al., 2019	http://www.nature.com/articles/s41467-019-10247-4
Software and algorithms		
RELION 3.1.3	Zivanov et al., 2018	https://github.com/3dem/relion
CryoSPARC v3.2	Punjani et al., 2017	https://cryosparc.com
MotionCor2	Zheng et al., 2017	https://msg.ucsf.edu/em/software/motioncor2.html
CTFFind4	Rohou and Grigorieff, 2015	https://grigoriefflab.janelia.org/ctf
Gautomatch	Kai Zhang	https://www2.mrc-lmb.cam.ac.uk/download/gautomatch-056/
Coot	Emsley et al., 2010	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/binaries/
Phenix	Liebschner et al., 2019	https://www.phenix-online.org/download/
UCSF Chimera	Pettersen et al., 2004	https://www.cgl.ucsf.edu/chimera/
PyMoL	Schrödinger	https://pymol.org/2/
Prism	GraphPad Software	https://www.graphpad.com/scientific-software/prism/
ImageJ	Schneider et al., 2012	https://imagej.nih.gov/ij/
SerialEM	David N. Mastronarde, 2005	https://bio3d.colorado.edu/SerialEM/