The interaction landscape between transcription factors and the nucleosome

Fangjie Zhu1, Lucas Farnung2, Eevi Kaasinen3, Biswajyoti Sahu4, Yimeng Yin3, Bei Wei1, Svetlana O. Dodonova2, Kazuhiro R. Nitta5, Ekaterina Morgunova3, Minna Taipale1,3, Patrick Cramer2,5 & Jussi Taipale1,3,4

Nucleosomes cover most of the genome and are thought to be displaced by transcription factors in regions that direct gene expression. However, the modes of interaction between transcription factors and nucleosomal DNA remain largely unknown. Here we systematically explore interactions between the nucleosome and 220 transcription factors representing diverse structural families. Consistent with earlier observations, we find that the majority of the studied transcription factors have less access to nucleosomal DNA than to free DNA. The motifs recovered from transcription factors bound to nucleosomal and free DNA are generally similar. However, steric hindrance and scaffolding by the nucleosome result in specific positioning and orientation of the motifs. Many transcription factors preferentially bind close to the end of nucleosomal DNA, or to periodic positions on the solvent-exposed side of the DNA. In addition, several transcription factors usually bind to nucleosomal DNA in a particular orientation. Some transcription factors specifically interact with DNA located at the dyad position at which only one DNA gyre is wound, whereas other transcription factors prefer sites spanning two DNA gyres and bind specifically to each of them. Our work reveals notable differences in the binding of transcription factors to free and nucleosomal DNA, and uncovers a diverse interaction landscape between transcription factors and the nucleosome.

The packaging of eukaryotic genomes is accomplished by histones, proteins that form an octameric complex that binds to the DNA backbone, forming nucleosomes1–4. In a canonical nucleosome, a 147-base pair (bp) segment of DNA is wrapped around the histone octamer in a left-handed, superhelical arrangement for a total of 1.65 turns, with the DNA helix entering and exiting the nucleosome from the same side of the histone octamer. The two DNA gyres are parallel to each other except at the position located between the entering and the exiting DNA, where a dyad region of approximately 15 bp contains only a single DNA gyre.

The nucleosome presents a barrier for the binding of proteins such as RNA polymerases to DNA5–8. Similarly, most transcription factors (TFs) are thought to be unable to bind to nucleosomal DNA9,10, except for a specific class of TFs called the pioneer factors11. Despite the importance of the nucleosome in both chromatin organization and transcriptional control12–17, the effect of nucleosomes on the binding of transcription factors has not been systematically characterized.

Nucleosome CAP–SELEX
To determine the effect of nucleosomes on TF–DNA binding, we developed nucleosome consecutive affinity purification–systematic evolution of ligands by exponential enrichment (NCAP–SELEX; Fig. 1a, Extended Data Fig. 1). The method is based on analysis of enrichment of specific sequences from complex 147-bp (lig147) or 200-bp (lig200) DNA libraries, containing 101- or 154-bp randomized regions, respectively. The sequences are reconstituted into a nucleosome, and the complexes incubated with TFs, which are subsequently purified and the bound DNA is recovered using PCR. After multiple selection rounds, the dissociated nucleosomal DNA is separated from intact nucleosomes. Analysis of the sequences enriched by NCAP–SELEX allows inference of TF binding specificities and positions on nucleosomal DNA, together with their effects on the stability of the nucleosome.

We performed SELEX both using nucleosomal (NCAP–SELEX) and free DNA (high-throughput SELEX8,19) using 413 human TF extended DNA binding domains (eDBDs) and 46 full-length constructs (Extended Data Fig. 1h, Supplementary Table 1). The selected TFs covered 29% of the high-confidence TFs from a previously published study20. The enriched sequences were analysed computationally using motif matching, de novo motif discovery and mutual-information (MI) pipelines (see Supplementary Methods). Because nucleosomes can affect TF motifs21, we primarily used a MI measure, which can capture any type of enriched sequence pattern (see Fig. 1b). Standard MI analysis also captures nucleosome sequence preference. To separate TF signals from the nucleosome signal, we limited the MI measure to the most highly enriched subsequences (enriched-sequence-based MI; E-MI; Fig. 1b). In parallel, we also analysed all data using motif-based approaches to explain and validate the findings (Supplementary Data 1, 2). Among the tested TFs, 220 eDBDs and 13 full-length constructs were successful (Fig. 1c; see Supplementary Methods for details).

Nucleosome inhibits TF binding
To determine the general effect of nucleosomes on TF–DNA binding, we analysed E-MI signals on lig200, which can accommodate only one nucleosome and contains both nucleosomal and free DNA (Fig. 2a, Extended Data Figs. 2, 3). On lig200 almost all TFs had a lower E-MI signal at the centre (Extended Data Fig. 2a), where the nucleosome occupancy is highest, indicating that the DNA-binding of most TFs is inhibited or spatially restricted by the presence of a nucleosome. However, the effect of the nucleosome on TF binding varied strongly between the TFs (Extended Data Fig. 2b, c). For example, SREBF2, RFX3 and JUND2 only show E-MI signal at the extreme ends of the ligand, suggesting that in the presence of free DNA, they are largely excluded from nucleosomal DNA. By contrast, other TFs...
Nucleosome CAP–SELEX. a, Schematic representation of NCAP–SELEX. The DNA ligands for SELEX contain a randomized region (grey) with fixed adaptors (blue). The protocol first selects ligands that are favourably bound by the nucleosome, and then from the nucleosome-bound ligand pool selects ligands that bind to a given TF. The orthogonal tagging of histone H2A (tag 1) and TFs (tag 2) enables the consecutive affinity purification. In the last (5th) cycle, the TF-bound DNA ligands are further separated into nucleosome-bound and unbound libraries before sequencing. b, TF-signal analysis by E-MI. Both the TF-binding signals (solid bar) and the nucleosome-binding signals (dotted bar) can be captured by the MI between 3-mer distributions at two non-overlapping positions of the ligand (left). In our analysis, we further focus on MI of the most enriched 3-mer pairs (E-MI, right) to filter out the nucleosome signals. Most analyses in this manuscript use the E-MI diagonal (box, containing E-MI from directly adjacent non-overlapping 3-mer pairs) because it contains the majority of TF binding signals and is generally similar to the motif-matching result (bottom). c, Family-wise coverage of successful TFs.

TFs can bind both nucleosomal DNA gyres

Some chromatin-modifying enzymes and synthetic molecules can bind both DNA gyres wrapped around the nucleosome. To explore whether TFs can also exhibit such a binding mode, we analysed the entire 2D E-MI signals. We found that binding of the T-box family TF brachyury (T) to nucleosomal DNA resulted in two prominent E-MI signals (Fig. 2b). One was located at the E-MI diagonal (that is, it was observed between adjacent subsequences), whereas the other resulted from sequences located approximately 80 bp from each other. The first signal represents binding of T to nucleosomal DNA similarly to free DNA. The second is associated with an approximately 80-bp motif, indicating dimeric binding that spans both DNA gyres (Fig. 2c). This type of binding was also observed for lig147 but not detected on free DNA (Extended Data Fig. 2e). The signal for the long motif is stronger on the ligands that remained bound to the nucleosome (Extended Data Fig. 2f), indicating that the gyre-spanning mode of T stabilizes nucleosomes. Similar binding was also observed for another T-box factor, TBX2 (Extended Data Fig. 2g), but not for other TFs. Despite the clear biochemical ability of T and TBX2 to bind to nucleosomal DNA using the cross-gyre motif, we did not identify this motif from available ChIP-seq data. Thus, the biological role, if any, of this binding mode needs to be addressed by further experimentation. For some TFs, we also identified weak signals for another binding mode, in which the TFs contact nucleosomal DNA at positions spaced approximately 40 bp apart (for example, TBX2 and ETV; Extended Data Fig. 2g). These results indicate that the nucleosome scaffold enables new binding modes for TFs that are not possible on free DNA.

Nucleosome affects the orientation of TF binding

In analysis of motif matches on lig200, we noted that the motifs of some TFs displayed a bias of matches in one orientation at the 5′ end, and in the other orientation at the 3′ end of the ligand. This pattern was observed for many ETS and CREB bZIP factors (Fig. 2d, e, Extended Data Fig. 4). The orientational preference induced by the nucleosome can be explained by the fact that nucleosome breaks the rotational symmetry of DNA (Extended Data Fig. 4d). Depending on TF orientation, a particular side of a TF will be in proximity with either the second gyre of nucleosomal DNA or the histone proteins.

To determine whether the directional binding of TFs to a nucleosome is also observed in vivo, we mapped nucleosome positions genome-wide in the human colorectal cancer cell line LoVo using micrococcal nuclease digestion with sequencing (MNase–seq). We found that the nucleosome distribution is asymmetric ($P < 0.0003$, two-sided t-test) around ELF1 and ELF2 in vivo sites (Fig. 2f, Extended Data Fig. 4e). Such asymmetry is not observed for the same ELF2 sites after salt treatment that laterally mobilizes the nucleosomes, or around ELF2 motif matches that do not show ChIP–seq signal (Fig. 2f). The nucleosome occupancy is lower upstream than downstream of the ELF2 sites. This pattern suggests that the more stable binding of ELF2 downstream of the nucleosome displaces the nucleosome or pushes it upstream. Several chromatin features that are asymmetric relative to sites occupied by TFs have been reported. Our observation that nucleosome itself induces asymmetry in the preferred binding orientation of TFs provides a potential mechanistic basis for these findings.
Fig. 2 | Nucleosome scaffolds DNA and breaks its rotational symmetry, enabling new TF binding modes. a. Schematic of single nucleosomes assembled on different positions of lig200, resulting in higher nucleosome occupancy towards the centre. b. Two different binding types of T (brachyury) on nucleosomal DNA. Heat map shows E-MI for all combinations of positions on lig200. Type 1 signal near the diagonal yields short motifs similar to those on free DNA. The type 2 signal corresponds to a motif approximately 80-bp long. Note that in contrast to type 1 signal, type 2 signal is not inhibited by the high nucleosome occupancy at the centre (arrowheads). c. Schematic of TFs (purple) that bind both gyres of nucleosomal DNA. d. Orientational asymmetry of binding of individual TFs on nucleosomal DNA. y axis: binding energy difference between two relative orientations of the most enriched subsequences. x axis: t-test P value of the difference compared to binding on free DNA (see Supplementary Methods). Note that most ETS-family TFs (red) show prominent asymmetry. Dot size represents the extent of signal enrichment in the NCAP–SELEX library of each TF. e. Orientational asymmetry of the ETS factor ELF2. At the 5’ end of the ligand, the ELF2 motif (top) is enriched on the minus strand, because ELF2 prefers to bind DNA in one orientation relative to the nucleosome (yellow, bottom left cartoon). At the 3’ end of the ligand, the ELF2 motif is enriched on the plus strand, as this leads to the same orientation of the ELF2 protein with respect to the nucleosome (yellow, bottom right). Note also that the two yellow ELF2 proteins make symmetric contacts, but to different strands of DNA (marked orange and purple; adapters are indicated in blue). Note that TF positions on the ligand are not fixed, for simplicity only a few example positions are shown. f. Asymmetric nucleosome distribution around genomic ELF2 sites (top, sites positioned at centre). Asymmetry is not observed for the same ELF2 sites after salt treatment to mobilize the nucleosome (bottom) or for ELF2 motifs without ChIP signal (middle). Nucleosome positions are shown as frequency of the centre of MNase fragments (140–170 bp). Each profile (n = 999 data points) is LOESS smoothed (locally weighted smoothing) with a span of 0.05 and the shaded band indicates the s.e.m.

Fig. 3 | Nucleosome induces positional preference to TF binding. a. Hierarchical clustering of the E-MI diagonals for NCAP–SELEX with the 147-bp ligand (lig147). E-MI diagonal is scaled for each TF (see Supplementary Methods). The names of the TFs are coloured by family. TFs from the same family tend to be clustered together. A few TFs were annotated as examples to illustrate their end (E), periodic (P) and dyad (D) preferences (see Supplementary Table 5). Note that the preferences are not mutually exclusive. Centre of the fixed position of nucleosome on lig147. b. E-MI penetration of each TF on lig147. All bZIP TFs are marked with red. c. E-MI diagonal and motif-matching results for the bZIP factor CEBPB. d. Schematic representation showing a TF that prefers the ends of nucleosomal DNA owing to breathing. Both ends of nucleosomal DNA will breathe but only one is illustrated here for clarity.
Nucleosome induces positional TF–binding preferences
Next we analysed the positional preference of TF binding to nucleosomal DNA. We designed the 147-bp NCAP–SELEX ligand (lig147) that matches the preferred length of nucleosomal DNA(29) allowing more precise mapping of TF-binding positions relative to the nucleosome. The results indicate that the presence of nucleosome restricts TF binding, and induces several types of positional preference (Fig. 3, Extended Data Figs. 5, 6). Expert analyses and machine learning analyses (Extended Data Fig. 6b, c, Supplementary Methods) revealed three types of positional preference on nucleosomal DNA (Fig. 3a, Supplementary Table 5): 1) end preference, these TFs prefer positions towards the end of the ligand that are partially accessible due to a process known as ‘breathing’(1,30,31). Many TFs of this class either radially cover more than 180° of the DNA circumference (for example, bZIP and bHLH), and/or bind to long motifs through a continuous interaction with DNA (for example, C2H2 zinc fingers) (Fig. 3a); 2) periodic preference, these TFs tend to bind to periodic positions on nucleosomal DNA and 3) dyad preference, these TFs prefer to bind to nucleosomal DNA near the dyad position.

Half of the circumference of nucleosomal DNA is in close proximity to the histones. As DNA is helical, equivalent positions that could be accessible to TFs are located at approximately 10-bp intervals. Accordingly, we found that many TFs prefer to bind to positions located approximately 10 bp apart on nucleosomal DNA (Fig. 3a, Extended Data Fig. 7). By applying a fast Fourier transform (FFT) to the E-MI diagonals, we obtained both the strength and rotational position (phase) of the approximately 10-bp periodicity for each TF (Fig. 4a). Analysis of the rotational position of binding for the TFs revealed that both major and minor grooves of nucleosomal DNA were accessible from the solvent side. For example, PITX and EOMES prefer almost opposite phases (Fig. 4a). This is consistent with the known structures; PITX contacts DNA principally via the major groove(32) (structure in Fig. 4b), whereas T-box TFs such as EOMES contact DNA mainly via the minor groove(33,34) (Extended Data Fig. 7b). Such periodic preference of binding has been reported previously for p53 and the glucocorticoid receptor(35,36), but the prevalence of this phenomenon was unclear. Among the TF families, periodic binding was particularly common among homeodomain TFs (Fig. 3a), and was also detected for homeodomain TFs from mouse liver (Extended Data Fig. 7g). Taken together, these results suggest that consistent with structural data(37) (Extended Data Fig. 5a), many TFs can bind nucleosomal DNA from the solvent-accessible side.

Analysis of the positional preference of TFs on nucleosomal DNA also revealed that the dyad region is strongly preferred by some TFs (Fig. 4c–g, Extended Data Fig. 8 and previous work)(38,39). For example, RFX5 shows very strong binding to the dyad positions of lig147 (Fig. 4c); on the basis of a competition assay, the affinity of RFX5 to dyad positions is higher than to free DNA (Fig. 4c, bottom; Extended Data Fig. 8b). To test whether RFX5 also prefers nucleosomal DNA in vivo, we expressed RFX5 in HEK-293 cells, and then detected nucleosome positions and RFX5-occupied sites using MNase–seq and MNase–ChIP. HEK-293 cells do not endogenously express RFX5, and in untransfected cells the positions at which exogenous RFX5 binds are located at a maximum of nucleosome occupancy (Fig. 4d, Extended Data Fig. 8). However, upon RFX5 expression, RFX5 forms a complex with nucleosomes, in which the positions of the nucleosomes are shifted to the sides of the sites that are bound by RFX5 (Fig. 4d, e). These results indicate that RFX5 prefers nucleosomal DNA in vivo, and that it potentially can induce nucleosome remodelling. In addition to RFX5, we also found that multiple SOX TFs have a preference for binding to dyad DNA (Fig. 4f, g). Such a preference was validated for SOX11 using an electrophoretic mobility shift assay (EMS; Extended Data Fig. 8). Taken together, our results indicate that on nucleosomal DNA, some TFs display a strong preference towards the dyad region.
Fig. 5 | Effects of TF binding on nucleosome stability. a, Hierarchical clustering of the differential E-MI diagonal between the nucleosome-bound and unbound cycle 5 libraries. Most TFs have stronger signal in the unbound library compared to that of the bound library, suggesting that their binding destabilizes the nucleosome. Brackets denote TFs that both destabilize and stabilize the nucleosome in a position-dependent way. Asterisks denote the ETS factors with a specific pattern of positional dependence. b, Mean strengths of E-MI diagonals in the nucleosome-bound and unbound cycle 5 libraries. The scatter plot shows the mean E-MI for the diagonals of each TF (dots), and for both the bound library (y axis) and the unbound library (x axis). The grey line represents where y = x. Most TFs have stronger signals in the unbound library (for example, CDX1, blue). A few TFs show the reverse (for example, TBX2, red, three replicates). For CDX1 and TBX2 the E-MI diagonals of the bound (B) and the unbound (UB) libraries are also illustrated. c, TF binding facilitates nucleosome dissociation. Binding of most TFs (magenta) to nucleosomal DNA leads to formation of a relatively unstable ternary complex (top right), and facilitates dissociation of the nucleosomes because the TFs prefer free DNA over nucleosomal DNA (bottom right). An alternative mechanism where the nucleosome first dissociates (bottom left) is not consistent with the observation that nucleosome induces positional binding preferences of TFs (see also Fig. 3). d, The identified major TF–nucleosome interaction modes.

**Effect of TF binding on nucleosome dissociation**

To determine whether TF binding affects the stability of the nucleosome, we performed an additional affinity capture step to separate the nucleosome-bound and dissociated DNA (unbound) in the last cycle of lig147 NCAP–SELEX (Figs. 1a, 5, Extended Data Fig. 9). Control experiments lacking TFs showed very little difference between the E-MI signal of the bound and unbound libraries, whereas in the presence of TFs, clear differences were observed (Fig. 5a, Extended Data Fig. 9a). We found that most TFs (for example, CDX1) have stronger E-MI in the unbound library compared to that of the bound library, suggesting that they can facilitate nucleosome dissociation upon binding (Fig. 5b,c). However, we also identified a few exceptional TFs, the binding of which stabilized the nucleosome. These include the T-box TFs, such as TBX2 (Fig. 5b). Moreover, the effect of TFs on nucleosome stability is also dependent on their binding mode and position on the nucleosomal DNA (Fig. 5a, Extended Data Fig. 9).

**Discussion**

TFs and the nucleosome are central elements regulating eukaryotic gene expression. In this study, we developed a new method, NCAP–SELEX, for analysis of nucleosome–TF interactions, and systematically examined the binding preference of 220 TFs on nucleosomal DNA. We identified five major interaction patterns between TFs and the nucleosome (Fig. 5d, Extended Data Fig. 10; Supplementary Table 5). The interaction modes are consistent with structural considerations, and not mutually exclusive. They include 1) binding that spans the two gyres of nucleosomal DNA; 2) orientational preference; 3) end preference; 4) periodic preference and 5) preferential binding to the dyad region.

Binding of most TFs facilitated the dissociation of nucleosomes. The simplest mechanism to explain this finding is that TFs bind to nucleosomal DNA and form a ternary complex. This complex is relatively unstable because the TFs prefer free DNA over nucleosomal DNA. This difference in affinity provides the free energy that facilitates dissociation of the nucleosome. Although the histone octamer binds 147-bp DNA more strongly than most TFs, within the approximately 10-bp segment that is bound by a TF, the bonds formed by the TF are stronger than those formed by histones. Therefore, binding of a TF to a partially dissociated nucleosome can also prevent rewinding of the TF-bound DNA segment to the nucleosome.

The TFs that facilitate the dissociation of nucleosome function as potential activators that can open chromatin and regulate gene expression. Some TFs, in turn, stabilized the nucleosome. These factors could repress gene expression, or to precisely position nucleosomes at specific genomic loci. Our findings are related to previous analyses that have identified pioneer TFs, which can access nucleosomal DNA. However, our observations indicate that a binary classification of TFs is not sufficient to capture the complete diversity of the interaction landscape between TFs and the nucleosome. Taken together, our results explain in part the complexity of the relationship between sequence and gene...
expression in eukaryotes, and provide a basis for future studies aimed at understanding transcriptional regulation based on biochemical principles.

Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Code availability
All of the computer programs and scripts used are either published or available upon request.

Data availability
All next-generation sequencing data have been deposited in the European Nucleotide Archive (ENA) under accession PRJEB22684. The relevant processed data are included as Supplementary Information.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0549-5.

Acknowledgements
We thank F. Zhong, A. Jolma, J. Zhang and J. Toivonen for valuable suggestions; E. Inns for proofreading; T. Kivosia for critical review of the manuscript and L. Hu, J. Liu and S. Augsten for technical assistance. This work was funded by the EU Horizon 2020 project MRGGrammar (664918), Cancerfonden (120529, 150662), Knut and Alice Wallenberg Foundation (2013.0088), Vetenskapsrådet (D0815201), Academy of Finland CoE (312042), (T.) DFG (SP8860, SPF1935), ERC Advanced Grant (639023), and EMBO fellowship ALTF 949-2016 (S.D.).

Reviewer information
Nature thanks T. Hughes, B. F. Pugh and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions
J.T., F.Z. and J.T. wrote the manuscript. All authors discussed the findings and contributed to the analysis of sequencing and structure data. F.Z. and J.T. interpreted the data and wrote the manuscript. All authors declared competing interests.

Competing interests
The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0549-5.

Supplementary information
is available for this paper at https://doi.org/10.1038/s41586-018-0549-5.

Reprints and permissions information
is available at http://www.nature.com/reprints.

Correspondence and requests for materials
should be addressed to J.T. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Extended Data Fig. 1 | See next page for caption.
Extended Data Fig. 1 | Experiment design and data analysis strategy of NCAP-SELEX. 

**a**, Expression of the recombinant histones from *Xenopus laevis*. For each lane 3 µg histone is loaded. Similar purifications for untagged H2A, H2B, H3, and H4 have been repeated at least three times. The SBP-H2A purification was performed once. 

**b**, Size-exclusion chromatogram of the histone octamer. Octamer formation was performed twice and the results were highly consistent. 

**c**, EMSA result showing the reconstituted nucleosomes using lig147 and lig200. The original ligands are also loaded as reference. The asterisks indicate the nucleosome bands. Similar results are seen in four independent nucleosome reconstitutions. For gel source data see Supplementary Fig. 1.

**d**, Oligonucleotide periodicity in the library enriched by nucleosome. As a quality control of nucleosome reconstitution, we verified whether nucleosome by itself is enriching the previously reported approximately 10-bp periodic oligonucleotide signal\(^41,42\). Nucleosome SELEX (without TF) were carried out for four cycles to enrich nucleosome-favouring ligands. The counts of each single and di-nucleotide across each individual ligand were Fourier transformed and summed up for the whole library. A clear peak around 0.1 bp\(^{-1}\) (corresponding to the reported approximately 10-bp periodicity) is visible for most mono- and dinucleotides. 

**e**, The C/G/CG preferences of nucleosome. All 9-mers were counted for the nucleosome-favoured (bound) and the nucleosome-disfavoured (unbound) libraries. The point representing each 9-mer is coloured according to its C/G/CG content (top), and the count ratios between the bound and the unbound libraries are summarized for 9-mers of different C/G/CG contents (bottom). For the box plots grouped by C/G content, the sample sizes of the boxes are 19,683, 59,049, 78,732, 61,236, 30,618, 10,206, 2,268, 324, 27 and 1, respectively for 9-mer groups containing 0 to 9 C/G. For the box plots grouped by CG dinucleotide content, the sample sizes of the boxes are 151,316, 91,824, 17,784, 1,200 and 20, respectively, for 9-mer groups containing 0 to 4 CG. The line within each box represents the median; the lower and upper boundaries of the box indicate the first and third quartiles and the whiskers represent the 1.5-fold interquartile range. More extreme values are indicated with dots. 

**f**, Analysis pipeline for the ligands enriched in NCAP-SELEX. 

**g**, E-MI strength comparison for libraries with and without TF signals. The E-MI heat maps represent signals in the input (cycle 0) library, in the cycle 4 library of nucleosome-favoured sequences (Nucl. SELEX), and in the NCAP- and high-throughput (HT)-SELEX cycle 4 libraries. The libraries enriched with TF (NCAP and HT) have much stronger E-MI signals compared to the cycle 0 and the nucleosome-SELEX library. The detected dimer signals of HSF1 in HT-SELEX is boxed. 

**h**, Family-wise coverage of TFs tried in NCAP–SELEX.

---

41. Collings, C. K., Fernandez, A. G., Pitschka, C. G., Hawkins, T. B. & Anderson, J. N. Oligonucleotide sequence motifs as nucleosome positioning signals. *PLoS ONE* **5**, e10933 (2010).
42. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).
Extended Data Fig. 2 | See next page for caption.
Extended Data Fig. 2 | NCAP–SELEX with lig200. a, Hierarchical clustering of the E-MI diagonals for NCAP–SELEX with the 200-bp ligand (lig200). The E-MI diagonal for each TF is oriented radially. The randomized region is 154 bp and contains 149 windows for MI calculation between neighbouring 3-mers. The names of the TFs are coloured by family with the colouring scheme indicated on the centre. TFs from the same family tend to be clustered together (for example, SOX, indicated). Because of the gradient of nucleosome occupancy, the penetration of the E-MI signal into the centre of the E-MI diagonals (E-MI penetration; see Supplementary Methods for details) reflects the ability of each TF to bind to nucleosomal DNA. Note that almost all TFs have lower E-MI towards the centre of lig200, indicating their lower affinity to nucleosomal DNA than to free DNA. The decrease of E-MI towards the centre is rarely observed in the absence of the nucleosome. Note that the binding inhibition of TF to nucleosomal DNA occurs in the absence of higher-order effects, such as chromatin compaction, remodelling or histone modification. This result directly verifies the mutually antagonistic role of TFs and the nucleosome, which has been biochemically validated in only a few cases.45,46. The E-MI diagonals shown are scaled for each TF (see Supplementary Methods). Owing to the fixed adaptor sequences, TFs may prefer one end of the lig200 over the other end. b, E-MI penetration of individual TFs on lig200. TFs are ordered according to their E-MI penetration depth towards the centre of the ligand. This order reflects the ability of TFs to bind nucleosome-occupied DNA. Note that the penetration of E-MI into the ligand centre (E-MI penetration; see Supplementary Methods for details) varies strongly between the TFs. TFs representing either of the two ends are coloured red and exemplified in c. The diagonal of E-MI for TFs with high (above dotted line) and low (below dotted line) E-MI penetrations. Because HT (blue) and NCAP–SELEX (black) may differ in stringency, each E-MI diagonal is normalized by dividing its maximum value. On lig200 the central 94 bp (shaded grey) is always occupied by a nucleosome. d, Correlation between E-MI penetration and the capability of TFs to bind nucleosomal DNA in vivo. Per base-pair coverage of MNase fragments (>140 bp) at ChIP–seq peaks of the TFs (x axis) is plotted against their E-MI penetration (y axis) in NCAP–SELEX. The calculation of Pearson’s r and the correlation test is performed for n = 20 TFs. The observed correlation suggests that the ability of TFs to bind nucleosomal DNA in NCAP–SELEX (E-MI penetration) partially explains the nucleosome occupancy at the sites of TFs in K562 cells. Thus the biochemical ability of TFs to bind to nucleosomal DNA also affects their binding in vivo. e, Left, E-MI heat map of T (brachyury) in HT-SELEX using lig200. Pairwise E-MI for all 3-mer pairs is presented as a heat map. The signal is only visible near the diagonal, no E-MI signal is detected across approximately 80 bp. Right, the gyre-spanning mode (arrow) of T (brachyury) on lig147. The corresponding motif is derived with the indicated seed for a specific position (number in the parentheses) in the high E-MI region (arrow). Position weight matrix generation follows our previous method47 using multinomial 1. f, Type 2 binding of Brachury (T) stabilizes nucleosome from dissociation. log2 ratio of E-MI between the bound and unbound libraries (cycle 5) is calculated for both the type 2 binding and for the background E-MI level (see Supplementary Methods for details) of Brachury (T). Compared to the unbound, the bound library has stronger type 2 binding but a similar background. As a control, for 20 random TFs (Rnd), the log2 ratio of E-MI between the bound and unbound libraries is also calculated for both the type 2 binding (hypothetic) and for the background E-MI level. For these TFs the bound libraries have similar E-MI strength as the unbound in the region corresponding to the type 2 binding of Brachury (T). Data are mean ± s.d.; two-sided t-test was used. 95% confidence intervals, 0.097 − 0.202 (T) and −0.008 − 0.004 (random TFs). The sample sizes are n = 20 libraries for random TFs and n = 4 independent SELEX replicates for Brachury (T). The raw data for the random control TFs are listed in Supplementary Data 3. g, E-MI heat map of TBX2, ETV4 and ETV1 in NCAP–SELEX using lig147. The E-MI signals across approximately 80 bp (type 2) or 40 bp (type 1) are indicated with arrows. The corresponding motif of each binding type is derived with the indicated seed for a specific position (number in the parentheses) in the high E-MI regions (arrows). Note that the E-MI signals across approximately 40 bp are position-specific, with one binding event being observed near the dyad, and the other(s) on the opposite side of the nucleosome, with the two contacts separated by approximately 180°. This binding mode can be achieved by TF dimers that contact nucleosomal DNA in a pincer-like manner. However, as the individual TFs are located far from each other in this binding mode, it probably suggests that the nucleosome may have two allosteric states, or may form a higher-order complex with these TFs.

43. Ramachandran, S. & Henikoff, S. Transcriptional regulators compete with nucleosomes post-replication. Cell 165, 580–592 (2016).
44. Li, M. et al. Dynamic regulation of transcription factors by nucleosome remodeling. Elife 4, e06249 (2015).
45. Sekiya, T., Muthurajan, U. M., Luger, K., Tulin, A. V. & Zaret, K. S. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. Genes Dev. 23, 804–809 (2009).
46. Hayes, J. J. & Wolfe, A. P. Histones H2A/H2B inhibit the interaction of transcription factor IIA with the Xenopus borealis somatic 5S RNA gene in a nucleosome. Proc. Natl Acad. Sci. USA 89, 1229–1233 (1992).
47. Jolma, A. et al. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res. 20, 861–873 (2010).
Extended Data Fig. 3 | Control experiments with lig200.

**a.** Determination of nucleosome positions for NCAP–SELEX libraries (lig200, all TFs). To examine if nucleosome has preferred positioning on lig200, nucleosomes were loaded onto the amplified cycle 4 NCAP–SELEX library of each TF. After digestion with MNase, the remaining DNA fragments were collected and sequenced. A titration was first carried out to find the appropriate concentration of MNase. As shown in the gel image (left, see Supplementary Fig. 1 for gel source image), 4.8, 2.4, 1.2, 0.6, 0.3, 0.15 U of MNase (lane 1–6) were added into each 25-μl reaction containing the purified nucleosome. According to the results, the condition marked by an asterisk was chosen for the reactions to determine nucleosome position. After sequencing, the fractions of MNase fragments that mapped to the variable region (grey) and to the adaptor-overlapping region (blue) of lig200 are visualized (middle, each row corresponds to a TF). To identify potential positional preference of nucleosome on lig200, the adaptor-overlapping fragments are analysed for their end distributions. Distributions of both the left end (cyan) and the right end (red) of the MNase-digested fragments on lig200 are shown (right, each row corresponds to a TF). Such distributions likely indicate that nucleosomes have two relatively preferred positions on lig200 (illustrated by cartoon in green). Note that most nucleosomes are not positioned by the adaptor (middle) thus are randomly distributed.

**b.** E-MI diagonals for HT-SELEX with the 200-bp ligand (lig200). TFs are arranged according to the clustering for NCAP-SELEX libraries (Extended Data Fig. 2a) to facilitate comparison. TFs without a lig200 HT-SELEX control are left as blank. The E-MI diagonal for each TF is oriented radially and the names of the TFs are coloured by family as indicated. The E-MI diagonals are scaled for each TF. Some TFs show preferred positions on lig200, probably due to the fixed adaptors.

**c.** TFs prefer free DNA to the edge of a nucleosome. For a few randomly chosen TFs, NCAP–SELEX was run using a ligand (Lig70N linker, sequence in Supplementary Table 2) that positions nucleosome at its centre by embedding a segment of Widom 601 sequence, and with randomized flankings. At a low resolution, the E-MI signal of TFs decreases monotonically towards the nucleosome-occupied region. Thus the higher E-MI at the flankings of lig200 (Extended Data Fig. 2a) suggests the preference of TFs for free DNA, rather than for the edge of a nucleosome. E-MI diagonals are scaled for each TF.

**d.** E-MI diagonals for TFs at doubled concentrations. The concentration effect on the E-MI diagonal of TFs is explored by running NCAP–SELEX at doubled (2 ×) concentrations for a few randomly chosen TFs. Compared to the E-MI diagonal with the original TF concentrations (1 ×), the change in the E-MI pattern is minor.
Extended Data Fig. 4 | See next page for caption.
Extended Data Fig. 4 | Nucleosome breaks the rotational symmetry of DNA. 

**a**, Density plot representing the orientational asymmetry of all TFs in NCAP–SELEX and in HT-SELEX. In NCAP–SELEX, more TFs bind with high orientational asymmetry than in HT-SELEX. A few TFs can also prefer different ends of the ligand for the two binding directions in HT-SELEX; this is likely induced by the adaptor sequences. However, there are more TFs with higher orientational asymmetry in NCAP–SELEX libraries, despite the fact that for most TFs their signals are stronger in HT-SELEX libraries. 

**b**, Orientation asymmetry of ELF2 revealed by using top 8-mers. Each row of the heat map corresponds to the counts distribution of a top 8-mer (non-palindromic) across the positions of the SELEX ligand. Hits of the top 8-mers occur at different ends for different strands of nucleosomal DNA (that is, an 8-mer and its reverse-complement prefer different ends), whereas their distribution is relatively homogeneous for free DNA. 

**c**, Orientation asymmetry of CREB TFs. CREB TFs have different motif density distributions for the two strands of nucleosomal DNA. The motif used for matching is indicated above. The minus strand profile is from the density of the reverse-complement motif. 

**d**, Break of the two-fold rotational symmetry of DNA induces preferred orientation of TFs. Left, free DNA has a pseudo-two-fold axis (red ellipse) perpendicular to the helix axis. Motifs in two orientations are symmetric with each other with respect to a 180° rotation centred on the axis. Right, for motifs on nucleosomal DNA, if the other strand of DNA or the histone proteins (green) affect binding, the two-fold axis of DNA no longer exists, as a 180° rotation centred on the axis no longer generates an identical conformation (the rotated image not superimposable with the original one). The break of rotational symmetry occurs also on the linker DNA that immediately flanks the nucleosome. 

**e**, Top, the orientational asymmetry of ELF1 in NCAP–SELEX of lig200. Bottom, the asymmetric nucleosome distribution around genomic ELF1 sites (top). Such asymmetry is not observed for the same ELF1 sites after a 30 min 500 mM KCl treatment to mobilize the nucleosome (bottom). ELF1 motif matches are positioned at the centre. Frequency of the centre of MNase-fragments (140–170 bp) is visualized for nearby regions to represent the nucleosome occupancy. Each profile (n = 999 data points) is LOESS smoothed with a span of 0.05 and the shaded band indicates the s.e.m. 

**f**, The orientational binding of ELF occurs on both the nucleosomal DNA and the nearby linker region. The motif matches of ELF on lig147 (top) suggest that the orientational binding occurs on nucleosomal DNA. In addition, the motif matches of ELF on the 293-hp ligand (bottom; nucleosome positioned at the centre, ligand schematic in Extended Data Fig. 3c) indicates that the orientational binding also occurs on nearby linker DNA regions.
Extended Data Fig. 5 | See next page for caption.
Extended Data Fig. 5 | TFs can bind nucleosomal DNA without substantial motif change. a, Cartoons showing that TFs are theoretically able to contact grooves of the bent nucleosomal DNA from the solvent-exposed side. The left panel for each TF shows the structures (Protein Data Bank (PDB); PITX2: 2LKX, TBX5: 2X6V). For the right panels of each TF, the PDB structure of the TF is aligned to the nucleosome structure (3UT9) as described in the Supplementary Methods (section 'FFT analysis and structure alignment'). The corresponding base pairs of the nucleosomal DNA were replaced with Coot48 according to the DNA sequence in the PDB structure of each TF. The models are visualized with UCSF Chimera49. b, TFAP binds nucleosomal DNA with slightly different specificity than free DNA. The scatter plot (top) shows the counts of gapped 9-mers from SELEX libraries of TFAP2B, enriched with NCAP–SELEX (x axis) and HT-SELEX (y axis). The examined 9-mers consist of three segments of trimers interspaced with two gaps (0–5 bp). Only the most enriched 9-mers (top 300 in each library and in the combined library) are shown for clarity. For comparison, the most differentially enriched gapped 9-mers were also used as seeds to derive the corresponding motifs from both libraries (right). The heat map (bottom) shows the pairwise E-MI for all combinations of positions on lig147, in the presence (left) and absence (right) of nucleosome. The arrowheads indicate the additional signals developed in the presence of nucleosome.

48. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. Acta Crystallogr. D 66, 486–501 (2010).
49. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612 (2004).
Extended Data Fig. 6 | See next page for caption.
Extended Data Fig. 6 | NCAP–SELEX with lig147. a, E-MI diagonals for HT-SELEX with the 147-bp ligand (lig147). TFs are arranged according to the clustering for NCAP–SELEX libraries (Fig. 3a) to facilitate comparison. The E-MI diagonal for each TF is oriented radially and scaled. The names of the TFs are coloured by family as indicated. b, The top five principal components (PCs) and the components from non-negative matrix factorization (NMF) with rank equal to five. The E-MI diagonals of lig147 (n = 195 TFs) were used in the dimension reduction. For visualization purposes, each component is centred and scaled. Note that the five principal components (left) correspond well to the three identified positional preferences of TFs on nucleosomal DNA (end: dim 1, 2; periodic: dim 3, 4; dyad: dim 5). c, Comparison between the scores from principal-component classifiers and custom classifiers. Red points indicate the TFs defined as displaying respective preferences according to custom classifiers. The PC classifiers are well in accordance with custom classifiers for the end and the dyad preferences (left), but not for the periodic preference (right). Because the phase of periodic preference can vary continuously whereas principal components can only capture discrete values, the custom FFT-based classifier is more natural for such purposes. The libraries of n = 195 TFs were used in the analyses. The correlation coefficients (Pearson’s r) are also indicated. d, E-MI diagonal and motif-matching results for the bZIP factor CEBPB. In HT-SELEX (without nucleosome), the binding signal is more distributed across the ligand. e, Pearson’s correlation between the E-MI penetrations of TFs on lig200 and on lig147. The libraries of n = 155 TFs, which are successful with both lig200 and lig147, were used in this analysis. The end preference of TFs on lig200 reveals that they prefer free DNA to nucleosomal DNA. The free-DNA preference also probably explains the end preference of TFs on lig147 owing to the observed correlation of E-MI penetrations. For each TF, the E-MI penetration values differ between lig147 and lig200 because free-DNA regions are expected near the ends of lig200, but not present on lig147. f, Correspondence between the E-MI patterns of TFs on lig147 and on lig200. The E-MI diagonals of RFX5 and SHOX on lig200 and those on lig147 are plotted together for comparison. The peaks on lig200 that illustrate the central preference of RFX5 and periodic preference of SHOX are indicated with red arrowheads. The weaker preference patterns on lig200 are due to the delocalization of the nucleosome on lig200, however they are still visible because the two fixed adaptors dictate two weakly preferred nucleosome positions.
Extended Data Fig. 7 | See next page for caption.
Extended Data Fig. 7 | TFs with periodic preferences. a, Density plot showing the periodicity strength of all TFs in NCAP–SELEX (orange) and HT-SELEX (blue). Note that the overall periodicity of E-MI is stronger for the NCAP–SELEX library compared to the free-DNA HT-SELEX library. b, A minor-groove binder prefers exposed minor grooves (m) on nucleosomal DNA. The E-MI diagonal of EOMES (T-box) is out of phase with the TA dinucleotide peaks, suggesting that it binds positions where the minor groove of nucleosomal DNA is facing outside (TA peaks indicate nucleosome–DNA contacts, whereas E-MI visualizes TF–DNA contacts, see Supplementary Methods for details). Accordingly, the TBX5 (T-box) structure (PDB entry 2X6V) shows contacts with DNA principally in the minor groove. Cartoon representation to the right shows that the steric hindrance is minimal when TBX5 (blue) binds out of phase with TA (orange) on the nucleosome structure (PDB entry 3UT9). c, Strength and phase of the approximately 10-bp periodicity of the TA dinucleotide in NCAP–SELEX and HT-SELEX libraries. For the library (lig147) enriched by a specific TF, the strength and phase information is derived from FFT of the TA counts at each position of the library. In the polar plot, each dot represents the library of one TF. The overall periodicity is stronger in the NCAP–SELEX libraries (yellow) than in the HT-SELEX libraries (blue), suggesting an enrichment of nucleosome signal. The TA phases in the NCAP–SELEX libraries of all TFs are similar, thus the rotational positioning of nucleosome on the SELEX ligand is similar for the libraries of all TFs. By contrast, the phase of the E-MI periodicity is much more dispersed (Fig. 4a), suggesting the preference of TFs towards different grooves of DNA. d, Cartoon representations of the 3D structures of PITX2 (PDB entry 2LKX) and TBX5 (T-box, PDB entry 2X6V) in complex with nucleosomal DNA. TBX5 structures were shown to illustrate the groove preferences of EOMES (T-box). The DNA ligand in the nucleosome structure (PDB entry 3UT9) contains phased TA steps (orange). Consistent with the SELEX result, PITX is more compatible with nucleosomal DNA when it binds in phase with TA, whereas T-box is more compatible when it binds out of phase with TA. Therefore, when a TF binds nucleosomal DNA according to the identified patterns, the steric conflict between TF and the histones is minimized. e, E-MI diagonal and motif-matching results for SHOX in NCAP–SELEX and HT-SELEX. The E-MI diagonal agrees with the motif-matching result. f, The approximately 10-bp periodicity for the preferred spacing of SHOX dimers on nucleosomal DNA. In NCAP–SELEX libraries of many periodic binders (SHOX as an example), enrichment of the most abundant 3-mer tandem repeats oscillates as a function of the spacing between the repeats. The enrichment is evaluated by the log2 ratio between the observed and expected occurrences. The observed approximately 10-bp periodicity with dimer spacing originates from the periodic availability of nucleosomal DNA. However, in most cases binding appears not to be cooperative, on the basis of the fact that the observed frequency of ligands with two motifs can be well estimated by the frequency of ligands that contain only one motif (data not shown). g, Homeodomain TFs from mouse liver prefer periodic positions on nucleosomal DNA. Motif hits of homeodomain TFs show a periodic pattern for both the nucleosome-bound and nucleosome-dissociated (unbound) libraries after incubation with mouse liver nuclear extract; however, the unbound library has more motif hits, indicating that binding events to the presented motif facilitate the dissociation of nucleosome. To more clearly visualize the approximately 10-bp periodicity, the Fourier-transformed spectra for both libraries are also shown to the right. The arrowhead indicates the peaks for the approximately 10-bp periodicity.
Extended Data Fig. 8 | See next page for caption.
Extended Data Fig. 8 | TFs with the dyad preference. a, E-MI diagonal and motif-matching results for RFX5. The distribution of binding events is more spread in the absence of nucleosome (HT-SELEX).

b, The design of the competition assay and the raw counts of RFX5 motif matches. Differently barcoded nucleosomal DNA (orange) and free DNA (blue) were mixed as input, and incubated with the TF protein. Purification for the TF-bound species was then performed. Matches of the indicated RFX5 motif was counted for both the nucleosomal DNA (orange) and the free DNA (blue), and for both of the input and the bound libraries. On nucleosomal DNA, more motif hits near the centre of the ligand are observed after purification. c, MNase–ChIP fragments near the binding sites of RFX5 and HOXB13. Motif matches within MNase–ChIP peaks of each TF are positioned at the centre. Counts of MNase–ChIP fragments are binned to 3 bp by 3 bp bins according to their lengths and centre positions. Nucleosome distribution is reflected by the signal intensity of the approximately 150-bp fragments (bracket). This visualization resembles the reported ‘V-plot’\textsuperscript{56}. Length distribution of all ChIP fragments and that of fragments <300 bp from the TF sites are shown on the right. Note that HOXB13 enriches ChIP fragments of approximately 120 bp at its sites (middle), suggesting that, similarly to most TFs\textsuperscript{50,51}, its binding sites in the genome are depleted of nucleosomes. By contrast, RFX5 enriches nucleosome-sized fragments (left). Most of the enriched fragments also have their centre positioned between the red ‘V’ lines, and thus overlap with the TF motifs. d, Nucleosome distribution near the binding sites of RFX5 and HOXB13 before transfection (no TF expression). MNase–seq fragments around the identified TF sites are visualized as in c. The sites later bound by exogenous RFX5 are located at the maximum of nucleosome occupancy (left). e, Nucleosome distribution near the binding sites of RFX5 and HOXB13 after transfection (with TF expression). The nucleosomes are now positioned beside the exogenous RFX5 sites (left). f, EMSA of SOX11 complexes with nucleosome and with free DNA. Nucleosome is reconstituted and purified using a modified Widom 601 sequence, which contains a SOX11 binding sequence (extracted from cycle 4 SELEX library) embedded close to the dyad. Each 40 μl reaction contains 1 μg DNA, together with SOX11 protein at a molar ratio of 0, 0.5, 1, 2 (indicated at the top of each lane) to DNA. Here the observed multiple shifts probably reveal the binding of SOX11 to additional weaker sites on the ligand (shown in g). For gel source data, see Supplementary Fig. 1. g, The score of SOX11 motif across the EMSA ligand (see Supplementary Methods for ligand sequence). The top three binding sites are indicated. h, DNA shape features around SOX11 motifs. DNA shape features were calculated using DNAshape\textsuperscript{R52,53}, for NCAP–SELEX (black), HT–SELEX (blue), and cycle 0 (input, grey) libraries. The black line is plotted last thus may hide other lines when all values are similar. The boundary of each motif is indicated with dashed vertical lines. Only the ligands with motifs around the centre (position range: 36–58) are included in the analysis.

50. Henikoff, J. G., Belsky, J. A., Krassovsky, K., MacAlpine, D. M. & Henikoff, S. Epigenome characterization at single base-pair resolution. Proc. Natl Acad. Sci. USA 108, 18318–18323 (2011).
51. Kasinathan, S., Orsi, G. A., Zentner, G. E., Ahmad, K. & Henikoff, S. High-resolution mapping of transcription factor binding sites on native chromatin. Nat. Methods 11, 203–209 (2014).
52. Chiu, T. P. et al. DNAshapeR: an R/Bioconductor package for DNA shape prediction and feature encoding. Bioinformatics 32, 1211–1213 (2016).
53. Chiu, T. P., Rao, S., Mann, R. S., Honig, B. & Rohs, R. Genome-wide prediction of minor-groove electrostatic potential enables biophysical modeling of protein-DNA binding. Nucleic Acids Res. 45, 12565–12576 (2017).
Extended Data Fig. 9 | See next page for caption.
Extended Data Fig. 9 | TF binding affects the stability of nucleosome.  

a, E-MI difference between the bound and the unbound cycle 5 libraries. The bound and the unbound libraries were collected either in the presence (left) or in the absence (right) of TFs. The heat maps visualize E-MI differences between the bound and unbound libraries for all position combinations of 3-mer pairs, and each pixel on the heat map is a mean of the E-MI difference of all the examined TFs at this pixel. For individual TFs, the value at each pixel is calculated as log₂(E-MI_{unbound}/E-MI_{bound}). Testing nucleosome dissociation in the absence of the TF aimed to verify whether the TF motifs on lig147 by themselves can affect the stability of the nucleosome. Note that in general, binding events close to the centre of nucleosomal DNA more efficiently dissociated the nucleosome (left). This observation is in accordance with the mutually exclusive nature between TFs and the nucleosome. Although TFs generally have lower affinity to the centre of the lig147, it is also conceivable that TF binding close to the centre will more efficiently undermine the DNA–histone interactions, and in turn lead to a higher rate of nucleosome dissociation. TFs bound close to the ends could have decreased the flexibility of the DNA there and subsequently disfavour the dissociation of DNA ends from the histones, which in turn contributes to nucleosome stability.  
b, The efficiency of nucleosome dissociation induced by ETV1 is dependent on its binding specificity. To displace nucleosome, binding with the shorter motif is more efficient than binding with the longer motif, because the shorter motif is more enriched in the dissociated library (unbound).  
c, Differential E-MI diagonals for TFs at doubled concentrations. The ability of each TF to dissociate or stabilize nucleosome is revealed by the log ratio of E-MI between the unbound and the bound cycle 5 libraries (differential E-MI). The concentration effect on the differential E-MI diagonal of TFs is explored by running NCAP–SELEX followed by the dissociation assay at doubled (2\times) concentrations of the TFs. The differential E-MI diagonals at 2\times TF concentrations resemble those at the original (1\times) TF concentrations.  
d, Differential E-MI diagonals for the four ETS family TFs indicated by asterisks in Fig. 5a.
Extended Data Fig. 10 | See next page for caption.
Extended Data Fig. 10 | Modes of TF–nucleosome interaction. a, For each TF, the strengths of all identified TF–nucleosome interaction modes, together with its ability to dissociate nucleosome, are shown in the heat map. The displayed features include the positional preference of each TF (E, end; P, periodic; D, dyad) on nucleosomal DNA, gyre-spanning binding mode (Gs), orientational asymmetry (Asym), and the ability of each TF to dissociate nucleosomes (Ds). TFs succeeding only in NCAP–SELEX with lig200 are presented to the right for their orientational asymmetry. In the heat map values are scaled into 0 to 1 for each mode, except for the dissociation, in which TFs that stabilize nucleosome are given negative values (green). The raw data are provided in Supplementary Table 5. b, All the identified modes can be explained by the structural features of nucleosome. TFs with the end preference (E) bind nucleosomal DNA close to the entry and exit positions. This preference is in line with the probability of spontaneous dissociation (breathing) of nucleosomal DNA, which decreases from the end to the centre54–56. TFs with a strong end preference are likely less compatible with nucleosomal DNA thus only bind to the dissociated regions. These TFs could be structurally hindered by nucleosome, because one side of the nucleosomal DNA is masked by the histones. Moreover, nucleosomal DNA is bent sharply, which could impair TF–DNA contacts if TFs have evolved to specifically bind to free DNA. TFs with the periodic preference (P) binds approximately every 10.2 bp positions on nucleosomal DNA. This preference arises also because nucleosomal DNA is accessible only from one side, which leads to significant accessibility change along each pitch (approximately 10.2 bp) of the DNA helix. TFs that bind to short motifs, or to discontinuous motifs, are still able to occupy the available periodic positions on nucleosomal DNA. TFs with the dyad preference (D) tend to bind close to the nucleosomal dyad. Structurally, the dyad is distinct from other regions of the nucleosomal DNA. The dyad contains only a single DNA gyre, and features the thinnest histone disk29,57. These characteristics of the dyad DNA reduce the steric barrier for TF binding. The relatively weak DNA–histone interaction around the dyad could allow TFs that bend DNA upon binding (for example, SOXs57) to deform DNA more easily at the dyad compared to other positions. In addition, the entry and exit of nucleosomal DNA are also close to the dyad; together with the dyad DNA, they provide a scaffold for specific configurations of TFs. FOXA has been suggested to make use of this scaffold to achieve highly specific positioning close to the dyad29,58. However, the dyad positioning of FOXA is not observed in this study using eDBD, potentially because the full length of FOXA is required for its interaction with the nucleosome59. A few T-box TFs were found to bind nucleosomal DNA with the gyre-spanning binding mode (Gs). This mode is observed because DNA grooves align across the two nucleosomal DNA gyres59. The parallel gyres could specifically associate with TF dimers, or TFs with long recognition helices or multiple DNA-binding domains. The dual-gyre binding is possible only on nucleosomal DNA, and it thus stabilizes the nucleosome from dissociation, and may therefore function to lock a nucleosome in place at a specific position. Many TFs such as ETS and CREB show an orientational asymmetry (Asym) upon binding to the nucleosomal DNA. The nucleosomal environment has induced such preference by breaking the local rotational symmetry of DNA. In accordance with the mutually exclusive nature of TF and nucleosome binding, most TFs were found to dissociate nucleosomes (Ds). While nucleosome weakens the affinity of incompatible TFs, binding of such TFs are expected to weaken the nucleosome–DNA contacts as well. The ability of TFs to dissociate nucleosome is required for them to open chromatin and to activate transcription. Moreover, we also observed TFs that both stabilize and destabilize nucleosomal DNA, depending on their relative position of binding. Such ability could be used to more precisely position local nucleosomes. All the identified TF–nucleosome interactions suggest that the TF–nucleosome interaction could be more complicated than the previously suggested pioneer/non-pioneer classification of TFs41. We observed that for eDBD of almost all TFs, including known pioneer factors such as FOX and SOX, free DNA was nonetheless preferred over nucleosomal DNA. However, some pioneer factors can bind relatively better to the interior of the nucleosome (for example, FOX and SOX). In addition, some other TFs prefer nucleosomal DNA at restricted positions, or with one of their multiple binding motifs. These strategies are likely related to the access of pioneer factors to nucleosomal DNA.

54. Polach, K. J. & Widom, J. Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. J. Mol. Biol. 254, 130–149 (1995).
55. Anderson, J. D. & Widom, J. Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. J. Mol. Biol. 296, 979–987 (2000).
56. Li, G., Levitus, M., Bustamante, C. & Widom, J. Rapid spontaneous accessibility of nucleosomal DNA. Nat. Struct. Mol. Biol. 12, 46–53 (2005).
57. Privalov, P. L., Dragan, A. I. & Crane-Robinson, C. The cost of DNA bending. Trends Biochem. Sci. 34, 464–470 (2009).
58. Ye, Z. et al. Genome-wide analysis reveals positional-nucleosome-oriented binding pattern of pioneer factor FOXA1. Nucleic Acids Res. 44, 7540–7554 (2016).
59. Cirillo, L. A. et al. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol. Cell 9, 279–289 (2002).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  - Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

- Clearly defined error bars
- State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

No software was used for Data collection

Data analysis

The softwares used in this manuscript include MOODS (v1.9.1), MACS (v2.1.1), Trim Galore (v0.4.3), samtools (v1.3.1), bwa (v0.7.16a), BEDtools (v2.26.0), Autoseed (Jolma, A. et al., Nature 527, 384-388 (2015); Nitta, K. R. et al, Elife 4, e04837 (2015)), PEAR (v0.9.8), UCSF Chimera (1.11.2), R (v3.4.0)

The R packages used include circlize (v0.4.3), ggplot2 (v2.2.1), FactoMineR (v1.39), NMF (v0.20.6)

Custom codes are available on request, this is stated in "Code availability" of the Methods section

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data
Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All next generation sequencing data are deposited to European Nucleotide Archive (ENA) under Accession PRJEB22684 and mentioned in Data Availability Statement. Most of our analyses are based on the E-MI diagonal values, the raw data is provided in Suppl. Table S3.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences

Study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | The sequencing depth of NCAP-SELEX and HT-SELEX libraries was set to ensure that at least hundreds of thousands unique reads are available for each TF. Under this sample size, if a TF is binding nucleosomal DNA without restrictions, any non-random pattern of TF binding that has a biologically meaningful effect size (as observed in our study) can only occur with an extremely small p-value. |
| Data exclusions | The failed SELEX experiments were excluded according to the QC criteria. The criteria define successful TFs as having detectably stronger E-MI between neighboring 3-mer pairs than that between 3-mer pairs far away from each other, and showing enriched motifs that are not contaminations from unrelated TFs (see "Quality control of the SELEX experiments" section in Methods). The exclusion criteria is established before we perform conclusion-related analyses. |
| Replication | We performed multiple cycles (4–5) of SELEX for each TF. Each cycle is essentially a replicate of the same experiment. In addition, the whole SELEX procedure was also repeated for all TFs. For all the reported signals, their enrichment is observed across multiple SELEX cycles, and are reproducible between two or more independent batches of SELEX. Two replicates are available for the in vivo validation by MNase-ChIP. |
| Randomization | Samples were analyzed directly and individually, and not randomized to experimental groups |
| Blinding | Most analyses were performed using computational algorithms. Investigators were not blinded. |

Materials & experimental systems

Policy information about availability of materials

Involved in the study

| n/a | Involved in the study |
| ☒ | Unique materials |
| ☒ | Antibodies |
| ☒ | Eukaryotic cell lines |
| ☒ | Research animals |
| ☒ | Human research participants |

Antibodies

| Antibodies | The following antibody was used in this experiment: V5 Tag Monoclonal Antibody from Thermo Fisher Scientific, catalog #: R960-25, RRID AB_2556564. The antibody was used at 1:200 dilution. |

Validation

Validated in previous publications for Human and Mouse, and by the following techniques: Immunocytochemistry (ICC), Immunoprecipitation (IP), Western Blot (WB), Immunohistochemistry (IHC), ELISA, Flow Cytometry, ChIP assay (ChIP), Immunohistochemistry (Paraffin) (IHC (P)).
Eukaryotic cell lines

Policy information about cell lines
Cell line source(s) LoVo (ATCC CCL-229), HEK293 (ATCC CRL-1573)
Authentication The LoVo and HEK293 cells were directly obtained from the trusted vendor (ATCC) and not from other laboratories and used within short time.
Mycoplasma contamination Tested to be free of mycoplasma infection by Hoechst staining
Commonly misidentified lines (See ICLAC register) No commonly misidentified cell lines were used.

Method-specific reporting

n/a Involved in the study
☒ ChiP-seq
☐ Flow cytometry
☐ Magnetic resonance imaging

ChiP-seq

Data deposition
☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.
☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

Files in database submission
Available in the project description of the linked database

Genome browser session
Available on request

Methodology

Replicates Two technical replicates are available for each MNase-ChIP experiment. Peaks in replicates enrich similar motifs.

Sequencing depth Total reads: 51285876 (RFX5) and 53017344 (HOXB13). Uniquely mapped reads: 9184303 (RFX5) and 9715845 (HOXB13). The sequencing run is paired-end. The sequencing length is 91 bp.

Antibodies V5 Tag antibody (R96025, ThermoFisher). All information available on vendor website

Peak calling parameters Peak called using MACS with
-f BAMPE -g hs -b -q 0.1 --nomodel -m 2 50
for MNase-ChIP__cell-HEK293__TF-RFX5__Replicate-1.bam and MNase-ChIP__cell-HEK293__TF-HOXB13__Replicate-1.bam

Data quality Correct TF motifs are discovered from the ChIP peaks. 6060 (HOXB13) and 1450 (RFX5) peaks are at FDR 5% and above 5-fold enrichment.

Software

bwa [Li H. and Durbin R., Bioinformatics 25, 1754-1760 (2009)]
MACS [Zhang et al., Genome Biol. 9, pp. R137 (2008)]