Systematic Study of Nucleosome-Displacing Factors in Budding Yeast

Graphical Abstract

Strong NDFs

We developed an assay to identify nucleosome-displacing factors (NDFs).

104 TFs are divided into three groups with strong, weak, and no NDF activities.

Fast and stable DNA binding underlies nucleosome-displacing activity.

DNA replication may facilitate NDFs to invade into nucleosomes.

Highlights

Authors
Chao Yan, Hengye Chen, Lu Bai

Correspondence
lub15@psu.edu

In Brief
Nucleosome-displacing factors (NDFs) open chromosomes and allow transcription factors to bind and initiate gene expression. Yan et al. developed a high-throughput method to identify NDFs and study their activities. Comparison between NDFs and other factors with no NDF activities generates insights into the mechanism of nucleosome invasion.
Systematic Study of Nucleosome-Displacing Factors in Budding Yeast

Chao Yan,1,2 Hengye Chen,1,2 and Lu Bai1,2,3,4,*
1Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA
2Center for Eukaryotic Gene Regulation, The Pennsylvania State University, University Park, PA 16802, USA
3Department of Physics, The Pennsylvania State University, University Park, PA 16802, USA
4Lead Contact
*Correspondence: lub15@psu.edu
https://doi.org/10.1016/j.molcel.2018.06.017

SUMMARY

Nucleosomes present a barrier for the binding of most transcription factors (TFs). However, special TFs known as nucleosome-displacing factors (NDFs) can access embedded sites and cause the depletion of the local nucleosomes as well as repositioning of the neighboring nucleosomes. Here, we developed a novel high-throughput method in yeast to identify NDFs among 104 TFs and systematically characterized the impact of orientation, affinity, location, and copy number of their binding motifs on the nucleosome occupancy. Using this assay, we identified 29 NDF motifs and divided the nuclear TFs into three groups with strong, weak, and no nucleosome-displacing activities. Further studies revealed that tight DNA binding is the key property that underlies NDF activity, and the NDFs may partially rely on the DNA replication to compete with nucleosome. Overall, our study presents a framework to functionally characterize NDFs and elucidate the mechanism of nucleosome invasion.

INTRODUCTION

Eukaryotic chromatin is densely packed with nucleosomes, which present a barrier for the binding of most transcription factors (TFs) (Adams and Workman, 1995). To allow the access of these TFs, regulatory regions in the eukaryotic genomes tend to be nucleosome depleted (Kaplan et al., 2009; Lee et al., 2007; Mavrich et al., 2008; Yuan et al., 2005; Zhang et al., 2009). The formation of the nucleosome-depleted regions (NDRs) is partly due to the mechanical properties of some DNA sequences (e.g., poly(A/T)) that directly lower the intrinsic stability of nucleosomes (Kaplan et al., 2009; Zhang et al., 2009, 2011). In addition, in budding yeast, a special group of TFs, referred to as “nucleosome-displacing factors” (NDFs), can antagonize nucleosome formation near their binding sites (Bai et al., 2011; Hartley and Madhani, 2009; Struhl and Segal, 2013; van Bakel et al., 2013). NDF was first proposed as the TF that establishes the nucleosome hypersensitive site on the GAL1-10 promoter (Fedor et al., 1988), and over the years, a few NDFs were discovered, including Abf1, Reb1, and Rap1 (Badis et al., 2008; Hartley and Madhani, 2009; Yarragudi et al., 2004). In multi-cellular eukaryotes, similar functions are carried out by pioneer factors (PFs), which invade into compact chromosomes and increase its accessibility (Hsu et al., 2015; Iwafuchi-Doi et al., 2016; Iwafuchi-Doi and Zaret, 2014; Zaret and Carroll, 2011).

By generating accessible regions on the chromosomes, NDFs and PFs allow other TFs to bind and initiate transcription. For example, on the CLN2 promoter in yeast, NDFs generate a ∼300 bp NDR that enables the binding and activation by activator complex Swi4/Swi6 (Bai et al., 2011). In ER-expressing MCF7 breast cancer cells, the binding and regulation by ER require the presence of a PF, FoxA1 (Carroll et al., 2005; Hurtado et al., 2011; Lupien et al., 2008; Malik et al., 2010).

NDFs and PFs are particularly powerful transcription regulators because they can target many locations in the genome and affect the expression of a large number of genes. Some PFs, such as FoxA1, Oct3/4, and Sox2, can therefore serve as “master regulators” that drive cellular differentiation or reprogramming (Lee et al., 2005; Takahashi and Yamanaka, 2006). Mutations in these factors tend to have severe physiological consequences. The known NDFs in budding yeast, including Abf1, Reb1, and Rap1, are all essential for viability. Mutation or mis-regulation of PFs in mammalian cells were shown to cause various types of cancers and developmental diseases (Lee et al., 2005; Magnani et al., 2015; Nakshatri and Badve, 2007; Wang et al., 2007).

A long-standing question in this field is what molecular mechanism(s) NDFs or PFs use to access their nucleosome-embedded sites, a context from which most other TFs are excluded. A few ideas were proposed by comparing the properties of NDFs/PFs with other TFs. A study in yeast found 10–20 TFs that contribute significantly to NDR formation and suggested that these NDFs may function through enriched interactions with chromatin remodelers (Ozonov and van Nimwegen, 2013). However, the NDFs in this work were identified based on the correlation between TF binding and nucleosome depletion, and the causal relation needs to be further established. In addition, interaction with nucleosome remodelers is unlikely the only distinct property of NDFs because factors like Swi4/Swi6 have no NDF activity but can interact with remodeling complex and histone chaperone (Takahata et al., 2008; Taneda and Kikuchi, 2004). The PF activities were thought to be associated with special
DNA binding domains (DBDs). FoxA1 has a winged helix DBD that resembles linker histone, which may allow it to recognize and bind to nucleosomes (Cirillo et al., 2002). For Oct4, Sox2, and Klf4, it was proposed that their pioneer activities are related to the ability to target partial motifs on the nucleosome surface (Soufi et al., 2015). Since these studies focus on a small number of factors, it is not clear how general these structural features apply to other PFs or NDFs.

To resolve these issues, we need a more complete list of NDFs/PFs so that we can systematically compare their properties with other TFs. In addition, an investigation of how binding site configurations, such as location and copy number, regulate the NDF activities can also shed light on the molecular process of nucleosome invasion.

In this work, we developed a novel high-throughput assay in yeast to identify NDFs among a large number of TFs and systematically evaluated the impact of the orientation, affinity, location, and copy number of their binding motifs on the nucleosome displacement. In this assay, we generated a complex library of synthetic regulatory elements, integrated them into the yeast genome, and measured the effect of these elements on the local nucleosome occupancy using micrococcal nuclease (MNase) digestion followed by massive parallel sequencing. The results from this assay allowed us to categorize 104 yeast TFs into groups 1, 2, and 3 with strong, weak, or no nucleosome-displacing activities, respectively. At least for some of these NDFs, their nucleosome displacing activities on the synthetic sequences agree well with their functions in the native genome. We showed that the NDF activities are highly dependent on the affinity of the motifs, but not on their orientation or locations within the nucleosome core particle. By comparing the properties of the three groups of TFs, we found that group 1 NDFs are distinguished from the other two by having high affinity and high concentration. A group 2 NDF can be converted into group 1 by overexpression, which displaces nucleosome in a cell-cycle-dependent manner. Together, these data lead to a replication-dependent kinetic competition model where NDFs bind to the transiently exposed binding sites after the replication fork and prevent nucleosome formation over the same DNA.

**RESULT**

**A Synthetic Biology Approach to Identify and Characterize NDFs in Budding Yeast**

To evaluate the nucleosome-displacing activity of a TF, the general idea is to engineer the binding site(s) of the TF into an otherwise nucleosomal sequence and measure the resulting changes in the nucleosome occupancy. We chose the HO promoter (HOpr) as the background sequence because it contains a well-positioned nucleosome array (from −1 to −7) (Bai et al., 2010; Takahata et al., 2009; Zhang et al., 2013). On the wild-type HOpr, the occupancies of these nucleosomes are dynamically controlled by a cell-cycle regulated activator Swi6 to allow the main activator, Swi4/Swi6, to bind and activate HO transcription (Bai et al., 2010; Takahata et al., 2009; Zhang et al., 2013). In this study, we used a modified HOpr with mutated Swi6 binding sites (swi6mut-HOpr) so that the nucleosome array becomes constitutive across the cell cycle (Figure 1A; Takahata et al., 2009; Zhang et al., 2013). The swi6mut-HOpr has essentially no transcriptional activity (Figure S1). When we engineered the binding site of a known NDF, Reb1 (Harley and Madhani, 2009; van Bakel et al., 2013), into nucleosome −4 at 33 bp away from the dyad (−33), this region became highly sensitive to MNase digestion (Figure 1B). The increased MNase accessibility indicates the loss of histone-DNA contacts near the Reb1 site, which we refer to as nucleosome displacement. In contrast, a defective Reb1 motif with 2 bp mutation at the same location has no such effect (Figure 1B). Similar nucleosome displacement can be seen with a Cbf1 consensus placed at the same location, but not with Mbp1 (Figure 1C). Reb1 inserted into a well-positioned nucleosome in the GAL10 ORF also generates a local NDR, indicating that the NDF activity of Reb1 does not require the swi6mut-HOpr context (Figure S2). HOpr variants containing a Reb1 or Cbf1 binding site fire stochastically in a cell-cycle-dependent manner (Figures 1D and S1), indicating that these NDFs enable the binding and activation by Swi4/Swi6. These results show that we can use this method to identify NDFs and quantify their nucleosome-displacing activities.

**Extension of the Synthetic Biology Approach to a High-Throughput Scale**

We next extended this method to a high-throughput scale to identify NDFs among a large number of TFs and systematically evaluated the effect of binding site configurations on their nucleosome-displacing activities (Figure 2A). Using the same nucleosome −4 sequence as the background, we designed and synthesized 16,667 types of DNA oligonucleotides containing various configurations of TF motifs (see STAR Methods). These oligonucleotides were divided into several sets, each aiming to test one aspect of the binding site configurations (i.e., type, orientation, affinity, location, copy number, and spacing) (Tables S1 and S2). Mutated motifs were also incorporated in the oligonucleotides as negative controls. These oligonucleotides were ligated into the swi6mut-HOpr on a plasmid, where they replaced the original sequence in nucleosome −4 from −45 to +15 (see STAR Methods). The complex plasmid library was linearized as a pool and integrated at a fixed genomic location (CLN2) into a yeast strain lacking the endogenous HOpr. We performed MNase digestion on the mixed yeast culture followed by amplicon sequencing over nucleosome −4. Oligonucleotide sequences depleted of nucleosome will be digested by MNase and thus generate low level of PCR products and sequencing reads, while sequences protected by nucleosome will survive MNase digestion and yield higher level of sequencing reads. Accordingly, we use the ratio of the sequencing reads from the MNase-treated sample to the undigested genomic DNA sample (total input) as a measure of nucleosome occupancy. For the motifs that decrease the local nucleosome occupancy, we define the corresponding factors as NDFs.

The pooled nucleosome occupancies on different sequences show a tailed distribution, with ~10% of them causing significantly reduced occupancy (Figure 2B). These measurements are highly reproducible among several biological replicates (Figure 2C). In addition, nucleosome occupancies measured by low-throughput method are consistent with the high-throughput data (Figure 2D). Together, we conclude that our method can
Figure 1. A Synthetic Biology Approach to Identify and Characterize NDFs in Budding Yeast

(A) Binding sites (rectangles) and nucleosomes (ovals) on the swi5<sup>mut</sup>-HOpr, which serves as the background (bkg) sequence in this study.

(B) Effect on the nucleosome occupancy by an engineered Reb1 binding site at 33 bp upstream the nucleosome –4 dyad (counted from the left edge of the Reb1 consensus). The lower panel shows the nucleosome occupancy measured by the MNase-qPCR assay in the region of nucleosome –5 to –3, and the x axis represents the mid-positions of the qPCR products relative to the ORF. The red arrow points to the location of the Reb1 motif (same as in C). Decreased nucleosome occupancy was observed with the consensus Reb1 site but not with the one containing a 2 bp mutation (Reb1mut). The error bars are the SE from three biological replicates.

(C) The same as in (B) except with different TF binding sites. The motif of Cbf1, but not of Mbp1, causes local nucleosome displacement.

(D) Representative traces of GFP intensity as a function of time in single cells measured by time-lapse fluorescence microscopy. GFP is driven by the bkg (gray) or the one with the Reb1 site at –33 (red). The latter shows cell-cycle-dependent GFP expression. Vertical dash lines mark the timing of cell divisions in the red trace.

generate robust and accurate nucleosome occupancy measurements over tens of thousands of integrated synthetic sequences.

**Group 1 NDFs: Strong NDFs that Displace Nucleosome through a Single Binding Motif**

There are six motifs that can lead to nucleosome –4 displacement when incorporated as a single copy, and their mutated versions have no effect on nucleosome occupancy (Figure 3A). The corresponding factors will be referred to as “group 1 NDFs.” Out of the group 1 NDFs, Abf1, Rap1, and Reb1 are “general regulatory factors” that are known to function as NDFs, which further verifies our results. The nucleosome displacing activity of Mcm1 was also implicated in previous studies (Bai et al., 2011; van Bakel et al., 2013). Consistent with their roles as NDFs, large fraction of the Abf1, Rap1, Reb1, and Mcm1 motifs in the native genome are exposed in the NDRs (82%, 48%, 83%, and 63%, respectively), much higher than the fraction for random sites (17%). Placing the motifs of group 1 NDFs on the Watson or the Crick strand does not significantly affect the nucleosome displacement level with the exception of Orc1, which shows a mild strand bias (Figure 3B).

Cbf1 was proposed to be a NDF in more ancient yeast species, but lost this activity in S. cerevisiae (Tsankov et al., 2011). We found that Cbf1 can still displace nucleosomes in S. cerevisiae, but only in the presence of high-affinity motifs (ΔΔG < 1.5 kcal/mole) (Figure 3C; Maerkl and Quake, 2007). Consistent with this result, ~86% of the high-affinity Cbf1 motifs in the yeast genome are exposed in NDRs, in comparison with 43% of the ones with lower affinity (Figure S3A). In addition, sites with medium affinity (1.5 < ΔΔG < 2.5 kcal/mole) show significant Cbf1 binding when exposed in NDRs but exhibit little binding when embedded under the nucleosome (Figure S3B; Zhou and O’Shea, 2011). Overall, these data indicate that Cbf1 and nucleosomes are mutually exclusive, and the high-affinity motifs are required for Cbf1 to displace nucleosomes.

Orc1 is part of the origin recognition complex (ORC) that binds autonomously replicating sequences (ARSs). ARSs are nucleosome-depleted in the genome, and they were thought to result from the polyT-rich ARS sequences (Eaton et al., 2010). In our oligo design, both the consensus and mutated Orc1 motifs contain polyT-track and cause similar level of intrinsic nucleosome instability (see STAR Methods), but the consensus motif can further reduce the nucleosome occupancy (Figure 3D). The same trend was not observed for other polyT-containing motifs, like Tho2 (Figure 3D). These data indicate that, besides the ARS sequence, the Orc1 protein (or other subunits in the ORC) has additional nucleosome-displacing activity. Consistent with this idea, inactivation of Orc1 leads to NDR shrinkage over ARSs in the genome (Eaton et al., 2010).

We next investigated the position dependence of the NDF activities by scanning the group 1 motifs from –42 to +1 across the nucleosome –4 with 1 bp step-size. Surprisingly, despite the different translational and rotational positioning on the nucleosome surface, Abf1, Rap1, Reb1, and the highest-affinity Cbf1
motifs displace nucleosome to a similar extent at all tested locations (Figure 3E). At least for Rap1, this result is in an apparent contradiction to the in vitro data that Rap1 association with a nucleosomal binding site is highly location dependent (Rossetti et al., 2001). These findings raise the possibility that, instead of invading into an intact nucleosome, these factors may bind when the nucleosome structure is transiently disrupted, e.g., during DNA replication (see later text for the test of this hypothesis). Motifs that displace nucleosome to a lesser extent, e.g., Mcm1 and Cbf1 with lower affinity, are slightly more effective when situated further away from the dyad (Figure 3E). In contrast, Orc1 site close to the dyad causes more reduction in the nucleosome occupancy, which is likely due to the lowered nucleosome intrinsic stability (Figure 3E).

**Group 2 NDFs: Weak NDFs that Displace Nucleosome through Multiple Binding Motifs**

In the genome, TF binding sites in the NDRs tend to cluster, suggesting that multiple TFs may work together to displace nucleosomes (Adams and Workman, 1995; Mirny, 2010). To test this idea, we incorporated 2 to 3 copies of the same motifs with 20 or 25 bp spacing in the oligonucleotide library. Group 1 NDFs can still displace nucleosome in this case, as expected (Figure 4A). More importantly, we found 21 other TF motifs that require multiple copies to generate nucleosome displacement ("group 2 NDFs;" Figure 4A). Spacing of 20 or 25 bp does not significantly affect the level of nucleosome displacement (Figure 4B).

Tbf1, a factor that was proposed to have NDF activities (Badis et al., 2008; van Bakel et al., 2013), turned out to be a group 2 NDF (Figure 4C). Rsc3, a subunit of the RSC chromatin remodel-
of Bas1 protein, the nucleosomes become well positioned with significantly increased nucleosome −4 occupancy, confirming that the Bas1 protein can influence the nucleosome positioning. This finding is in agreement with previous data that bas1 mutation leads to higher nucleosome occupancy in SHM2 and GCV2 promoters (with multiple Bas1 sites), but not in GID8 promoter (with single Bas1 site) (Zhu and Keeney, 2015). Ume6 is usually thought as a transcriptional repressor that decreases the accessibility of chromatin (Goldmark et al., 2000; Kadosh and Struhl, 1997). By comparing the nucleosome occupancy ± Ume6, we found that this factor has NDF activity that leads to the formation of a short NDR at its binding sites with two well-positioned nucleosomes in the immediate vicinity. There are also two cases where deletion of a group 2 factor, Rgt1 or Uga3, does not affect nucleosome occupancy (Figure S4). It is possible that Rgt1 and Uga3 are not NDFs and the nucleosome displacement is caused by other factors that recognize similar motifs. Alternatively, Rgt1 and Uga3 can displace nucleosomes, but their functions can be replaced by redundant factors upon their deletion. These possibilities require future investigation.

Multiple Rap1 Sites Generate Nucleosome-Sized Particles that Do Not Contain Histones

Previous studies have identified nucleosome-sized particles in the yeast genome with the degree of digestion highly sensitive to MNase concentration (Kubik et al., 2015). The nature of these particles remains controversial. Some studies attributed these particles to fragile nucleosomes (Kubik et al., 2015), while others...
proposed that they are non-histone protein complexes (Chereji et al., 2017). Our library sequences with engineered binding sites allow us to investigate this issue in a more controlled manner.

We repeated the high-throughput measurements in Figure 2 with different MNase concentrations (Figure S5A). We found a distinct subset in which the measured occupancy has a strong dependence on the MNase level (Figure 5A). Many sequences in this subset contain 2–3 copies of Rap1 motifs (Figures 5A, 5B, and S5B), a protein that was found to associate with the MNase-sensitive particles in the native genome (Kubik et al., 2015). Interestingly, sequences with a single Rap1 motif do not have such a property (Figures 5A and 5B). Using MNase-histone ChIP, we did not detect histone H3 in a MNase-sensitive particle generated by 3 Rap1 sites (Figure 5C; see STAR Methods), indicating that the partial protection from MNase is provided by Rap1 and maybe other non-histone proteins instead of fragile nucleosomes. It is likely that multiple Rap1 proteins are required to generate a sufficiently large footprint to be detected in our assay. It is yet to be seen whether the MNase-sensitive particles in the native genome are generated by the same mechanism.

Differences among the Three Groups of TFs

Besides groups 1 and 2, there are many TFs (“group 3”) that are available in the nucleus, but their motifs do not show any nucleosome displacing activities in our assay (Figure 6A; see STAR Methods). Consistent with these observations, most group 3 motifs in the yeast genome are not enriched in the NDRs (Table S3). The NDR enrichment of a subset of group 3 motifs is likely due to their co-localization with other NDF binding sites (Bai et al., 2011).

We next investigated whether there are any systematic differences among the three groups of TFs. We looked into a number of biophysical and structural properties of these TFs and found that group 1 factors are distinguished from the other two groups...
with higher binding selectivity and higher concentration (Figure 6B; see STAR Methods). To test whether these properties are sufficient in generating NDF activity, we overexpressed a group 2 factor with high binding selectivity, Rfx1, to see whether it can be converted into group 1 at higher concentration. In the wild-type cells, nucleosome is fully occupied over a template containing a single Rfx1 motif (Figure 6C). At higher concentration, Rfx1 turns into a group 1 factor and causes local nucleosome displacement near its binding site (Figure 6C). Introducing a bacteria protein Tet repressor (TetR) into yeast also opens a nucleosome gap over its binding site (Figure 6C). These results indicate that rapid and stable binding between NDFs to DNA is important for their ability to displace nucleosomes. In addition, since NDF activities are concentration dependent, the list in Figure 6A may change under different growth conditions.

Group 2 and 3 TFs do not differ significantly in terms of binding selectivity or concentration (Figure 6B). According to the STRING database (Szklarczyk et al., 2017), group 2 TFs also do not have more interactions with histones, histone acetyltransferase, or nucleosome remodeling complexes (p value > 0.1; see STAR Methods; Table S4). Interestingly, 65% of the group 2 factors have zinc-cluster DBDs, and this fraction is significantly higher than those in group 1 (0%, p < 0.01) and group 3 (14%, p < 0.001) (Figure 6D). For a group 2 NDF (Ume6), its zinc-cluster DBD alone is sufficient in generating a local NDR over the Ume6 binding sites (Figure 6E; see STAR Methods), demonstrating that, even for group 2 NDFs, the DBD binding activity underlies nucleosome displacement. It is possible that the zinc-cluster structure provides certain advantage of recognizing or competing with nucleosome. The detailed molecular mechanism is not clear.

Cell-Cycle-Dependent Nucleosome Displacement

We next used the Rfx1 overexpression system to test our previous hypothesis of DNA-replication dependent nucleosome invasion. We used alpha-factor to generate G1-arrested cells (with regular cycling cells as control), induced Rfx1 overexpression for 2 hr, and measured the nucleosome occupancy before and after the induction (see STAR Methods). Rfx1 overexpression caused a partial NDR formation in the cycling cells but had no effect on the G1-arrested cells (Figure 7A). We also tested cell-cycle dependence of the nucleosome displacement by Cbf1. We induced Cbf1 in a cbf1Δ strain and measured nucleosome occupancy over the Cbf1 binding site in both cycling and G1-arrested cells (see STAR Methods). We observed nucleosome displacement in both cases, with a higher displacement rate in the cycling cells (Figures 7B and 7C). Overall, the data indicate that a cell cycle event (likely DNA replication) promotes nucleosome displacement by NDFs, and this effect may be more prominent for weaker NDFs. It also shows that there are factors, like Cbf1, that can invade into nucleosome even in the absence of cell cycle.

DISCUSSION

A Systematic Investigation of NDFs using Synthetic Sequence Library

The ability to control gene expression level is ultimately encoded in the enhancer and promoter sequences. Synthetic enhancer/promoter libraries have been employed to understand the “grammatical rules” of gene regulation, e.g., how copy number, orientation, affinity, and spacing of activators affect the expression level (Gertz et al., 2009; Sharon et al., 2012; Smith et al., 2013). Nucleosome occupancy on these synthetic sequences, however, was either not taken into account or artificially depleted through poly(A/T) tracks. A recent study investigated nucleosome displacement by NDFs using synthetic library, but included only 12 factors (Levo et al., 2017). To our knowledge, our study is the first large-scale analysis of NDF activities. This approach allows us to discover ~20 new NDF motifs. The same approach can be applied to higher eukaryotic species to identify and characterize PFs.

NDF/PF activities are traditionally measured by overexpressing or deleting a TF and probing the nucleosome occupancy change on the native genome (Badis et al., 2008; Hartley and Madhani, 2009; Magnani et al., 2011; Pataskar et al., 2016; van Bakel et al., 2013). In comparison to this method, our synthetic biology approach offers some unique and complementary advantages. First, the traditional method requires genetic manipulation of TFs one at a time and therefore suffer from low throughput. Our method significantly enhances the throughput by testing >10^4 sequences with different TF motifs arranged into different configurations simultaneously. Second, NDFs or PFs are often essential for cell viability and differentiation, and deletion of these factors can have global and pleiotropic effects. By manipulating binding sites instead of the TF itself, we essentially work with the wild-type cells and thereby avoid

Figure 5. Multiple Rap1 Sites Generate Nucleosome-Sized Particles that Do Not Contain Histones

(A) Nucleosome occupancies measured at high versus low MNase concentrations (5 versus 1.25 unit/mL). Each dot represents a particular library sequence. The dots in magenta are the sequences containing 2–3 Rap1 binding motifs; the yellow ones contain 1 Rap1 motif.

(B) Nucleosome occupancy over sequences containing 3 or 1 Rap1 motif measured with low (blue), medium (purple), or high (red) MNase concentration.

(C) MNase-H3 ChIP measurements over the sequence in (B) containing three Rap1 motifs with low (blue) or high (red) MNase concentration.

The error bars in (B) and (C) are the SE from three biological replicates.
the secondary effects. Third, the native NDRs are often bound by multiple NDFs. Depending on whether they function independently, cooperatively, or redundantly, deletion of a NDF may or may not lead to increased local nucleosome occupancy. In contrast, the method here is able to determine the causal relationship between factor binding and nucleosome occupancy change by working with a unique, well-defined background sequence. Finally, the same sequence with different chromosome context can have altered accessibility to NDFs or PFs (Garcia et al., 2010; Iwafuchi-Doi and Zaret, 2014), which adds...
ing motifs (Figure 3C); and Ume6 DBD alone is sufficient in generating Cbf1 functions as NDF only in the presence of high-affinity binding a NDR (Figure 6E). These observations indicate that fast and stable DNA binding plays a key role in promoting the NDF activity.

TF that binds DNA with high affinity, can displace nucleosome into group 1 through overexpression (Figure 6C); TetR, a bacterial abundance (Figure 6B); a group 2 NDF, Rfx1, can be converted NDFs distinguish themselves by having high affinity and high nucleosome invasion mechanism of yeast NDFs. First, group 1 members that allow them to bind nucleosome with high affinity (Soufi et al., 2015; Zaret and Carroll, 2011). It is yet to be seen whether Cbf1 or other NDFs identified here have similar properties. 

Mechanism of the Nucleosome Invasion
We collected two lines of evidence that provide insights into the nucleosome invasion mechanism of yeast NDFs. First, group 1 NDFs distinguish themselves by having high affinity and high abundance (Figure 6B); a group 2 NDF, Rfx1, can be converted into group 1 through overexpression (Figure 6C); TetR, a bacterial TF that binds DNA with high affinity, can displace nucleosome over its binding site when overexpressed in yeast (Figure 3C); Cbf1 functions as NDF only in the presence of high-affinity binding motifs (Figure 3C); and Ume6 DBD alone is sufficient in generating a NDR (Figure 6E). These observations indicate that fast and stable DNA binding plays a key role in promoting the NDF activity.

Second, for most group 1 NDFs, the level of nucleosome displacement does not have strong dependence on the positions of their binding site within the nucleosome from 40 to the dyad (Figure 3E); overexpressed Rfx1 displace nucleosome in free-cycling but not G1-arrested cells (Figure 7A); and Cbf1 can displace nucleosome in both cycling and G1-arrested cells, and the former happens at a faster rate (Figures 7B and 7C). These results indicate that NDFs may take advantage of certain cell cycle event to gain access to free DNA. As these measurements were performed on only limited number of factors, the generality of these conclusions requires further testing.

Based on the evidence above, we propose that there are two ways NDFs can invade into nucleosomes: replication dependent or replication independent. In the first scenario, NDFs bind to the transiently exposed sites right after the replication fork and prevent nucleosome formation over the same DNA. For this mechanism to work, it is essential for NDFs to have high concentration to compete effectively with the rapid replication-dependent nucleosome assembly, and also to have long dwell time on the DNA (i.e., high affinity) to prevent subsequent nucleosome assembly. This mechanism does not require the NDFs to directly interact with histones, nucleosome remodeling machinery, or other nucleosome-related factors. It also does not require NDFs to have special DBDs. In principle, any abundant TFs that bind tightly to DNA, including artificially introduced bacterial TFs, can displace nucleosome through this mechanism. Such relaxed requirement implies that many TFs can potentially serve as NDFs, consistent with the lack of homology and conservation among NDFs: group 1 factors all have diverse DBDs, and some of them are not conserved among different yeast species (Wapinski et al., 2007). This replication-dependent mechanism may work particularly well in a fast-growing species like yeast where replication occurs frequently.

The Cbf1 results in Figure 7 indicate that, besides replication-dependent invasion, some NDFs can also directly target and compete with nucleosomes in the absence of replication. Such behavior of Cbf1 makes it more similar to PFs in higher eukaryotes, which are thought to directly target nucleosome-embedded sites and can change DNA accessibility even in G1-arrested cells (Donaghey et al., 2018; Iwafuchi-Doi and Zaret, 2014). PF competition with assembled nucleosomes after replication is also supported by a recent study in Drosophila, which showed that many enhancer NDRs are lost after replication and that it takes hours for the NDRs to reappear and TFs to re-bind (Ramachandran and Henikoff, 2016). Such direct invasion is likely to be critical for cellular reprogramming in terminally differentiated cells (Hanna et al., 2008). The property of Cbf1 that enables its direct invasion is not clear. For PFs, it was proposed that the pioneer activities rely on special DBD structures that allow them to bind nucleosome with high affinity (Soufi et al., 2015; Zaret and Carroll, 2011). It is yet to be seen whether Cbf1 or other NDFs identified here have similar properties.

NDF Coordination with Remodeling Factors to Determine the NDR Size
Following their initial binding, NDFs cause subsequent changes in the local chromatin structure. Interestingly, in a few cases where we mapped the nucleosome occupancy over a 400 bp
region near the synthetic NDF motifs, we found very different nucleosome positioning patterns: Reb1, Cbf1, and Rap1 generate wide NDRs that expand over 100 bp, while Ume6 generates a narrow one that resembles a linker DNA (Figures 1B, 1C, 4E, and 5B). Bas1 causes fuzzy positioning of the neighboring nucleosomes without forming a well-defined NDR (Figure 4E). Since the mapped regions have identical DNA sequences except for the NDR binding sites, the difference in nucleosome positioning likely reflects the differential NDF activities.

TFs can recruit remodeling enzymes to slide or evict nucleosomes (Boeger et al., 2003; Svaren and Horz, 1997). Also, different sets of remodelers are targeted to different genes even though their binding to DNA lack sequence specificity (Yen et al., 2012). We therefore hypothesize that NDFs coordinate with different remodelers to shape the local chromatin structure. Indeed, Reb1 was found to recruit RSC, a remodeling enzyme in the same family as Swi/Snf (Hartley and Madhani, 2009). The remodelers from this family was found to “push” nucleosome away from the recruitment point (Dechassa et al., 2010), leading to the formation of a broad NDR. In contrast, Ume6 recruits ISW2, which “drags” nucleosome toward itself and produces a narrow NDR (Stockdale et al., 2006). Ume6 DBD alone can generate a NDR longer than the full-length Ume6 but similar to TetR (Figure S6), likely because the Ume6 DBD can no longer interact with ISW2. The patterns of nucleosome positioning around Reb1 and Ume6 are consistent with their respective roles as general regulatory factor and repressor. Apart from these well-studied cases, the functional coordination between most NDFs and remodeling factors is not clear, and future studies are needed to systematically map these relations.

**REFERENCES**


## STAR METHODS

### KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibodies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antibodies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histone H3 C-terminal Antibody</td>
<td>Epicypher</td>
<td>Cat# 13-0001</td>
</tr>
<tr>
<td><strong>Bacterial and Virus Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEB 5-alpha Competent <em>E. coli</em> (High Efficiency)</td>
<td>New England Biolabs</td>
<td>Cat# C2987</td>
</tr>
<tr>
<td>NEB 5-alpha Electrocompetent <em>E. coli</em></td>
<td>New England Biolabs</td>
<td>Cat# C2989</td>
</tr>
<tr>
<td><strong>Chemicals, Peptides, and Recombinant Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nuclease micrococcal from Staphylococcus aureus</td>
<td>Sigma</td>
<td>Cat# N3755</td>
</tr>
<tr>
<td>D-(-)-Galactose</td>
<td>Sigma</td>
<td>Cat# G0750</td>
</tr>
<tr>
<td>D-(-)-Raffinose pentahydrate</td>
<td>Sigma</td>
<td>Cat# 83400</td>
</tr>
<tr>
<td>Protease Inhibitor Cocktail</td>
<td>Sigma</td>
<td>Cat# P8215</td>
</tr>
<tr>
<td>Protein A/G PLUS-Agarose</td>
<td>Santa Cruz Biotechnology</td>
<td>Cat# sc-2003</td>
</tr>
<tr>
<td>Proteinase K</td>
<td>VWR</td>
<td>Cat# 97062-670</td>
</tr>
<tr>
<td>Z1005 Zymolyase</td>
<td>United States Biological</td>
<td>Cat# 37340-57-1</td>
</tr>
<tr>
<td><strong>Experimental Models: Organisms/Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yLB91</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY32-Ume6</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY32-Bas1</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY32-Rgt1</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY32-Uga3</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY32-Mbp1</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY32-Reb1</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY32-1mut</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY32-Cbf1</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY57-Ume6</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY57-Bas1</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY57-Rgt1</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY57-Uga3</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY50</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY55-Rfx1</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY55-tetR</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY62-Rap1</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY62-3Rap1</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY61-Rfx1</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY61-TetR</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY63-Rfx1</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY65</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY66</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td>yCY67</td>
<td>This study (Table S5)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Oligonucleotides</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Library Amplification Primers</td>
<td>This study (Table S6)</td>
<td>N/A</td>
</tr>
<tr>
<td>Designed oligo library</td>
<td>This study (Table S3)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Recombinant DNA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pCY15</td>
<td>This study</td>
<td>N/A</td>
</tr>
<tr>
<td>pCY20</td>
<td>This study</td>
<td>N/A</td>
</tr>
</tbody>
</table>

(Continued on next page)
CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to the Lead Contact, Lu Bai (lub15@psu.edu).

METHOD DETAILS

Plasmids and yeast strains

Standard methods were used to construct the strains and plasmids. See Table S5 for plasmid and strain list. The background strain for the yeast library, yLB91, was derived from W303 with part of the endogenous HO promoter URS2 (−1078 to −281 relative to HO ORF) deleted to avoid complication in the subsequent PCR step. Modified HOpr driving destabilized GFP (Bai et al., 2010) was cloned into the multiple cloning site.

Library design and construction

We retrieved from ScerTF a collection of 196 sequence-specific factor binding motifs, their position weight matrix (PWM), and recommended PWM cutoff (Spivak and Stormo, 2012). We also chose 225 Cbf1 motifs with different affinities from previous in vitro measurement (Maerkl and Quake, 2007). For motifs that severely overlap with each other, we varied 1-2 “less” consensus bases to make them more unique while modified motifs still have PWM scores above the recommended threshold. We ranked the possibility for these motifs to be NDF sites based on their enrichment in genome-wide NDR sequences (determined based on genome-wide nucleosome occupancy data from Lee et al. (Lee et al., 2007)) (Table S1).

The library oligonucleotides were designed using MATLAB programs developed in the lab. The final sequences are 100bp in length, comprised of central variable 60bp flanked by 20bp constant sequences on each side (used as primers) (Table S6). The central 60bp sequences are based on the native sequence in the HOpr/C0 4 nucleosome (from −45 to +14 bp relative to the dyad), and we replaced part of the sequences with various TF binding motifs at different positions. The oligonucleotides were designed as a few sets, and each set incorporated different factors (Table S1). For example, in the “single copy” and “multiple copy” sets, we included all known motifs; in the “positioning” set, we only incorporated motifs of known NDFs or the ones that are highly enriched in NDRs. A total of 16667 oligonucleotides were designed and commercially synthesized at Agilent (Table S2).

Both the synthetic oligonucleotides and the background vector pCY15 were engineered with two BbsI sites to generate compatible sticky ends. Because BbsI is a Type IIS enzyme, its recognition sequence will be eliminated upon digestion, leaving a sticky end that is part of the native sequence of nucleosome −4. To construct the plasmid library, we first PCR amplified the oligonucleotides, digested with BbsI enzyme and ligated to BbsI-digested pCY15 (Table S6). For each 10,000 types of oligonucleotides, we started with ~10ng single stranded DNA (~1E15 molecules) in a 2,000μl PCR reaction, ran for 12 cycles, and used an Amicon Ultra-2 10K centrifugal filter to exchange the PCR solution for 1X NEB buffer 2.1. We then adjusted the DNA concentration and proceeded with Bbsl digestion at 37°C overnight (Table S6). The 60bp band in the digestion product was gel-purified (Sun et al., 2012) and ligated to Bbsl-digested pCY15. A total of 150μl ligation was made and incubated overnight at 16°C. The ligation product was concentrated through a spin column into a 30μl volume and transformed into 5-alpha Electrocompetent E. coli (NEB). A total volume of 250μl electrocompetent cells transformed with 25μl ligation product typically produced ~2,000,000 colonies. Those cells were then pooled and subjected to a maxi prep.

To construct the yeast library, we digested the plasmid library in the CLN2 degron sequence and integrated into the yLB91 strain at the CLN2 locus through homologous recombination. We used an electroporation protocol (Lin et al., 2011) to generate ~300,000 transformed yeast colonies (on average ~20 colonies per oligonucleotide). The transformed yeast cells were pooled together for the nucleosome occupancy measurement.

Measurement of nucleosome occupancy

To evaluate the nucleosome occupancies over our synthetic sequences, we first performed the MNase assay as previously described (Bai et al., 2011) with four different MNase concentrations (0.625 / 1.25 / 2.5 / 5 unit/ml, Sigma N3755). We grew the yeast
library in synthetic complete media to OD_{660} 0.2, used 75ml of the culture for each MNase digestion reaction, and gel-purified the mono-nucleosome bands. We also purified undigested genomic DNA as the input control.

**Amplicon sequencing**

To sequence the integrated oligonucleotide sequences, we conjugated them to Illumina TruSeq adaptors and carried out amplicon sequencing on a NextSeq500 machine with 1 x 80bp single end sequencing (35% phiX spike-in for each flow cell). The amplicon was generated through two PCRs. The first PCR was carried out with 19 cycles in a total volume of 500µl (Table S6). We included 4 random bases right after the sequencing initiation base to allow a better cluster separation in the initial calibration step of Nextseq machine. The second PCR was carried out with 6 cycles in a total volume of 200µl (Table S6), which produced a 244bp amplicon ready for sequencing.

**Sequencing data analysis**

We mapped the raw sequencing reads to our designed oligo library through Burrows-Wheeler Aligner with default setting, and 80%–90% of the reads were mappable. The mapping yielded the sequencing read counts for each type of oligonucleotides in the MNase sample and the genomic DNA sample. The reads in the genomic sample needs to be larger than 100 for any oligonucleotide to be analyzed further. Out of the 16,667 designed oligonucleotides, 16,153 (97%) passed this threshold. The ratio of the read count in the MNase digested sample to the genomic sample was calculated for each oligonucleotide, and normalized based on the nucleosome occupancy on the background sequence (Table S2).

**Further selection of TF motifs**

During data analysis, we further inspected the TF motifs to generate the table in Figure 6A. We removed all the factors that are expressed at undetectable level or not localized in the nucleus (Ghaemmaghami et al., 2003). These TFs are likely to be unavailable under our growth condition, and therefore we cannot comment on their NDF activities. We also eliminated the factors with ambiguous concentration / localization information. Among the rest, we removed a few TFs that bind indirectly to DNA, such as Gal80 (binds through Gal4) and Swi6 (binds through Swi4). These factors by themselves do not have sequence specificity and do not fit the definition of NDF. Finally, there are a few factors with overlapping motifs with known NDFs. For example, YDR026C has very similar motif as Reb1, and the nucleosome displacement over the YDR026C motif is likely caused by Reb1. These factors are also excluded from Figure 6A. For the full list of excluded factors and the nucleosome occupancy over their motifs, see Table S7.

**Low throughput test on individual TF and template**

1) Rfx1/TetR overexpression: strains yCY61-Rfx1/TetR were constructed through a two-step transformation. The nucleosome −4 sequences containing Rfx1 or TetR binding motifs were integrated to CLN2 locus, and Met25 promoter driving RFX1 or TETR were integrated into the TRP1 locus. To measure the NDF activity with or without Rfx1/TetR overexpression, these strains were incubated overnight in synthetic complete media −/+ methionine and then subjected to MNase-qPCR assay. 2) Replication dependent test (Rfx1): We first constructed yCY63-Rfx1: a strain based on yCY61 but further with bar1 deletion strain, and measured the nucleosome occupancy over 3 sequences. 3) Cbf1 NDF activity in cycling / G1-arrested cells: we replaced the endogenous Cbf1 promoter with the DRB1 promoter in the deletion strain, and measured the nucleosome displacement over the YDR026C motif is likely caused by Cbf1. 4) We used the endogenous UME6pr driving UME6-DBD (727–836) (Anderson et al., 1995) in the ume6 deletion strain, and measured the nucleosome occupancy over 3 UME6 motifs. 5) Cbf1 NDF activity in cycling / G1-arrested cells: we replaced the endogenous Cbf1 promoter with GAL1 promoter in a bar1 background. The strain’s doubling time is approximately 3 hr in synthetic medium containing raffinose. We grew the cells to log phase in raffinose before treating them with alpha factor for 3 hr. We then added galactose to the medium to reach a final concentration of 3%. At 0, 0.5, 1, 2, and 3 hr, we collected the cells and performed MNase assay on those samples. Cycling cells (without alpha factor treatment) were induced for same time and used as a control. The degree of nucleosome displacement is quantified by nucleosome occupancy at −716 (normalized by the occupancy at −892).

**QUANTIFICATION AND STATISTICAL ANALYSIS**

**Model and Bioinformatics**

1) Calculation of the intrinsic nucleosome stability. We ran the model in Kaplan et al. (Kaplan et al., 2009) with the “Log-Ratio Binding mode” to compute the free energy of nucleosome formation on the synthetic 147-bp nucleosome −4 sequences. 2) Scanning the yeast genome for consensus binding sites: using the PWM and the recommended cutoff for each TF (Spivak and Stormo, 2012), we calculated the PWM score for every N-bp segment from the yeast genome (N is the length of the binding site). The segments...
with the score higher than the cutoff were considered as consensus. 3) The relative positioning of TF motifs and nucleosomes: we identified genome-wide NDRs as previously described (Bai et al., 2011). Motifs that fall into these regions were considered NDR-localized. From the rest, we selected the ones with local nucleosome density > 0 (log2) as motifs embedded under the nucleosomes. The others were ambiguous and excluded from the statistics. 4) Localization, concentration, and binding selectivity of TFs: The localization and concentration data are from Ghaemmaghami et al. (Ghaemmaghami et al., 2003). Undetectable factors, as well as factors that are not localized in the nucleus, were considered unavailable and excluded from the analysis in Figure 4. If a TF is localized in both nucleus and cytoplasm, it was assumed that it has equal concentration in the two compartments, and therefore the protein copy number in the nucleus was calculated as 7% of the total (nucleus volume is ~7% of the cell volume). The binding selectivity of TFs were determined statistically by their PWM (Ozonov and van Nimwegen, 2013). 5) Interaction between TF and other factors: we first used the GO terms from SGD to find genes encoding histones, subunits of acetylation enzymes, and subunits of remodeling complexes (Table S4). We then counted all the known interactions between these factors with each TF in Group 1, 2, and 3 based on the string database (Szklarczyk et al., 2017) (e.g., there are 4 known interactions between Abf1 and all the histones) (Table S4). There are no significant differences between the number of interactions found for Group 1, 2, and 3 TFs.