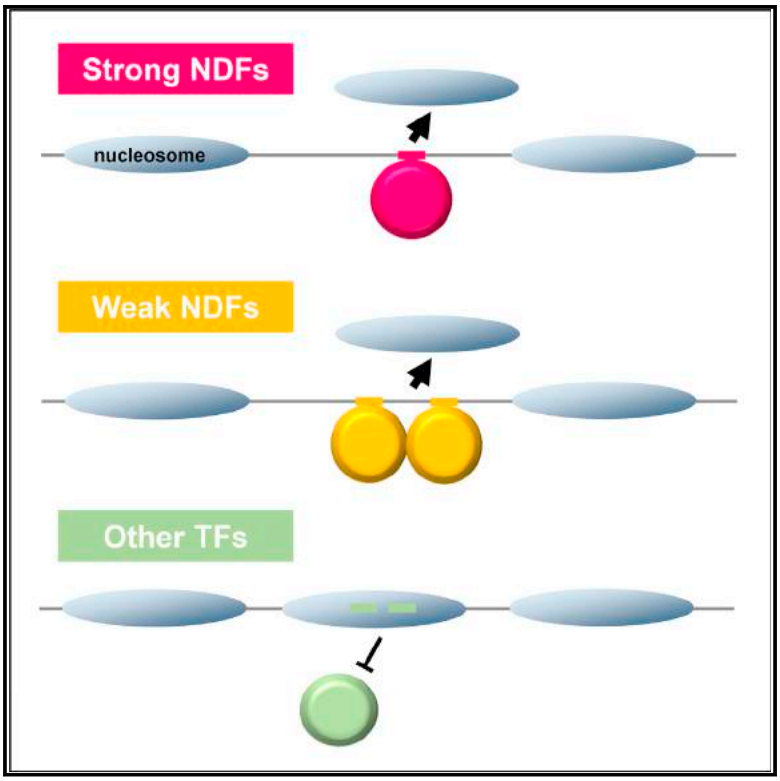
Article

Molecular Cell

Systematic Study of Nucleosome-Displacing Factors in Budding Yeast

Graphical Abstract



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.

Highlights

- 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

Yan et al., 2018, Molecular Cell *71*, 294–305
July 19, 2018
https://doi.org/10.1016/j.molcel.2018.06.017

CellPress

CellPress

Molecular Cell
Article

Systematic Study of Nucleosome-Displacing Factors in Budding Yeast

Chao Yan,[1,2] Hengye Chen,[1,2] and Lu Bai[1,2,3,4,*]
[1]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA
[2]Center for Eukaryotic Gene Regulation, The Pennsylvania State University, University Park, PA 16802, USA
[3]Department of Physics, The Pennsylvania State University, University Park, PA 16802, USA
[4]Lead 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 nuclease 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* promotor 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., 2009; 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 Swi5 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 Swi5 binding sites (*swi5^{mut}-HOpr*) so that the nucleosome array becomes constitutive across the cell cycle (Figure 1A; Takahata et al., 2009; Zhang et al., 2013). The *swi5^{mut}-HOpr* has essentially no transcriptional activity (Figure S1). When we engineered the binding site of a known NDF, Reb1 (Hartley 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 *swi5^{mut}-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 *swi5^{mut}-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

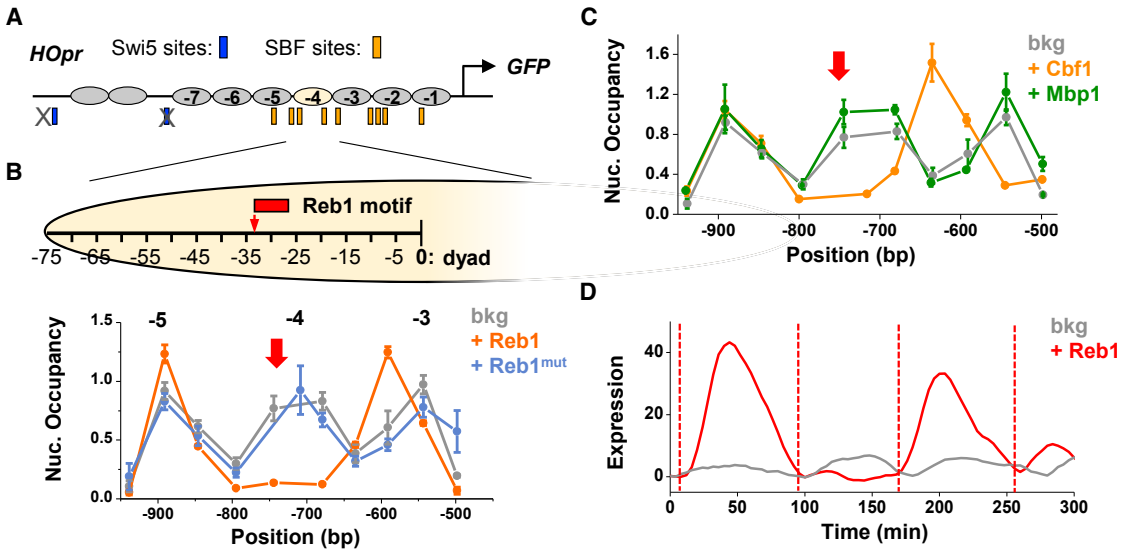


Figure 1. A Synthetic Biology Approach to Identify and Characterize NDFs in Budding Yeast

(A) Binding sites (rectangles) and nucleosomes (ovals) on the *swi5*^{mut}*-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 (Reb1^{mut}). 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 *HOpr* (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

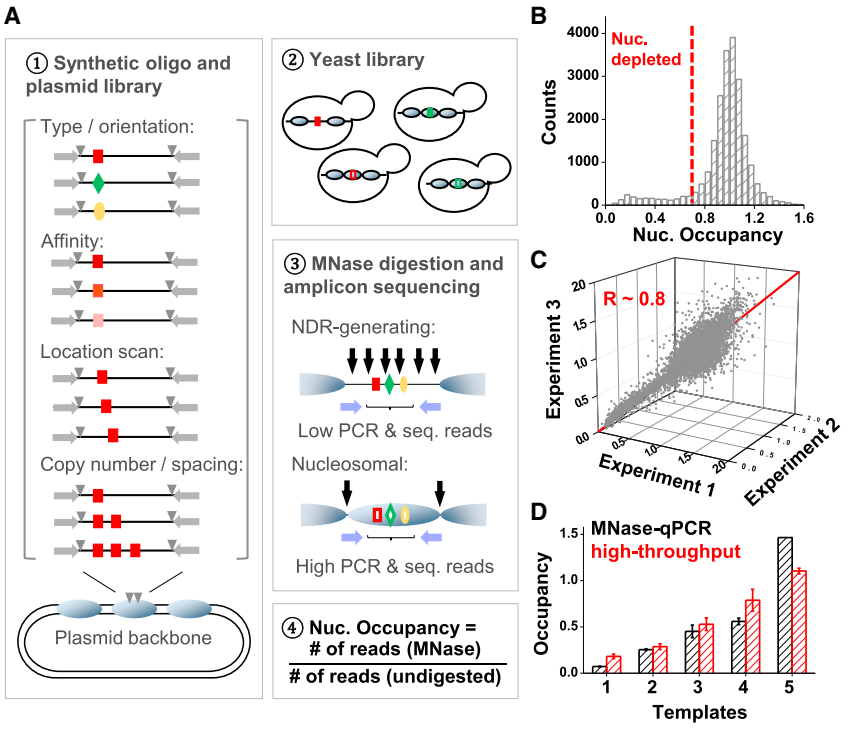


Figure 2. Extend the Synthetic Biology Approach to a High-Throughput Scale

(A) Procedure of the high-throughput method.

(B) Histogram of nucleosome occupancies on different oligo sequences integrated into the yeast genome. Occupancy 0.7 (3 standard deviations lower than the mean of the higher peak) is used as the threshold for significant nucleosome depletion.

(C) Comparison of the occupancy data from three biological replicates. The pairwise Pearson correlation coefficient between any two of the three datasets are 0.79, 0.81, and 0.84.

(D) Nucleosome −4 occupancies measured by low-throughput MNase-qPCR assay are consistent with the high-throughput assay. Template 1-5: 1: Reb1 at −43; 2: ORC1 at −43; 3: Ume6 at −40, −20, and 0; 4: Bas1 at −37, −17, and 3; 5: Ino2 at −37, −17, and 3. The error bars are the SE from three biological replicates.

motifs displace nucleosome −4 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 −4 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 remodeling complex that has sequence-specific DNA binding activity (Angus-Hill et al., 2001), also belongs to this category. To understand whether multiple Tbf1 or Rsc3 are required to displace nucleosome in the native genome, we examined the nucleosome occupancy over genome-wide Tbf1 and Rsc3 sites (van Bakel et al., 2013). Nucleosomes are significantly displaced over the Tbf1 clusters (multiple Tbf1 motifs located closely), but not over single Tbf1 motifs (Figure 4D). Such displacement is Tbf1 dependent, as Tbf1 inactivation significantly increases the nucleosome occupancy (Figure 4D). The nucleosome recovery is incomplete with Tbf1 inactivation, likely because NDRs on the native genome tend to contain multiple redundant NDFs, and mutation in one factor usually has limited effect on the nucleosome occupancy (Bai et al., 2011). A similar trend was also observed for Rsc3 (Figure 4D). For most group 2 TFs, their motif clusters tend to be much more enriched in the NDRs than single motifs (Table S3). In contrast, single motifs of group 1 factors Abf1 and Reb1 are sufficient to cause nucleosome displacement (Figure 4D). Therefore, group 1 and 2 NDFs also behave differently in their natural context. For Tbf1, the copy-number-dependent nucleosome displacement may have physiological significance. Most Tbf1 clusters locate in the subtelomeric regions where they function as "barriers" to prevent the spread of silencing, and nucleosome disruption was proposed to contribute to this function (Bi and Broach, 2001). In contrast, Tbf1 single sites tend to locate in the regular euchromatic regions where they can serve as enhancer blockage insulators, and this function is unrelated to nucleosome positioning (Yan et al., 2015). Therefore, Tbf1's function in the genome may be modulated by the variations in the copy number of its motifs.

Because our assay identifies NDF motifs, but not the factors, we deleted a few group 2 TFs to directly test their NDF activities. The results of Bas1 and Ume6 are shown in Figure 4E. In *BAS1*$^{+}$ cells, nucleosome positioning is rather "fuzzy" over the sequence containing three Bas1 binding sites. Upon the deletion

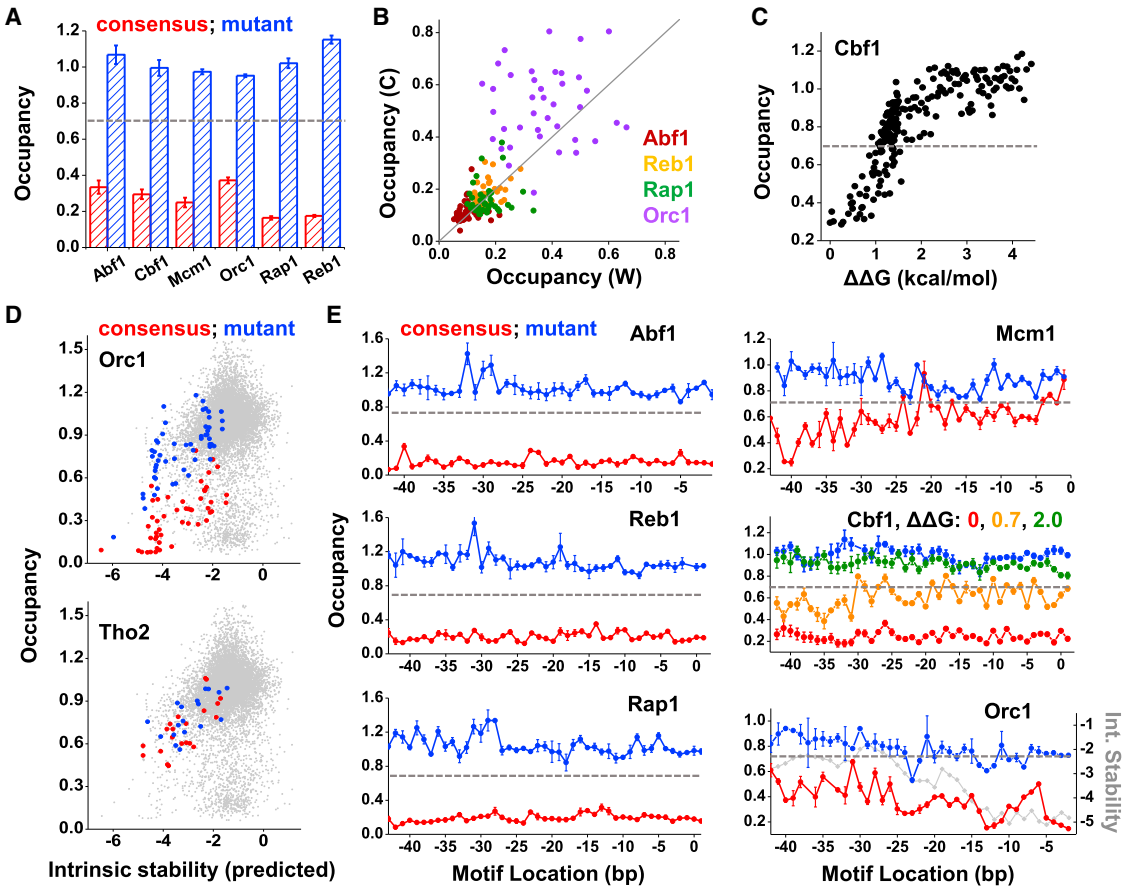


Figure 3. Group 1 NDFs: Strong NDFs that Displace Nucleosome through a Single Binding Motif

(A) For six factors, a single motif can decrease nucleosome −4 occupancy. The plot shows nucleosome occupancies in the presence of the consensus (red) or mutant (blue) motif of these factors at location −40 in nucleosome −4. The dashed line marks the threshold of nucleosome displacement (same for below).

(B) Nucleosome occupancy with the motifs on either the Watson (W) or Crick (C) strand. Only four factors were shown here because the other two, Cbf1 and Mcm1, have palindromic motifs. There are many data points for each factor, as their motifs were placed at variable positions in the nucleosome −4.

(C) Nucleosome occupancy in the presence of Cbf1 motifs with different binding affinities. In this plot, all Cbf1 motifs are located at −40. ΔΔG (Maerkl and Quake, 2007) is the Cbf1 binding affinity relative to the strongest binding site: larger ΔΔG represents weaker binding.

(D) Occupancy versus intrinsic stability of nucleosome −4. The gray background includes all the nucleosome −4 sequences in our library, and the red and blue dots highlighted the sequences containing consensus (red) or mutant (blue) motifs of Orc1 or Tho2.

(E) Nucleosome occupancy as a function of the location of group 1 TF motifs. Consensus (red) or mutant (blue) sites were scanned through the nucleosome with 1 bp interval from −43 to the dyad. Cbf1 motifs with three different affinities were shown here. The intrinsic stability of nucleosome −4 with the consensus Orc1 motif at different locations was also plotted (gray).

The error bars in (A) and (E) are the SE from three biological replicates.

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

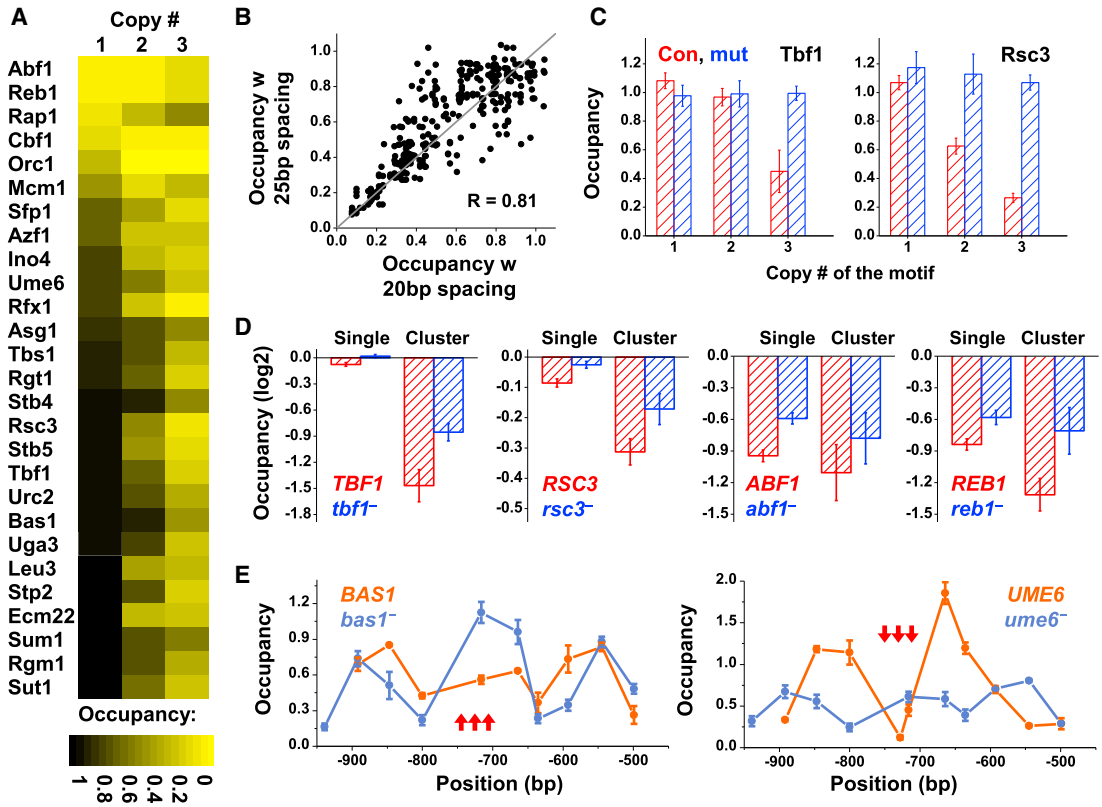


Figure 4. Group 2 NDFs: Weak NDFs that Displace Nucleosome through Multiple Binding Motifs

(A) Heatmap of nucleosome occupancy on sequences containing 1–3 copies of TF motifs. The lighter color represents lower nucleosome −4 occupancy. For each copy number, the motifs were placed at a few different positions, which result in similar level of nucleosome occupancy. The plot shows the average occupancy among these positions.

(B) The effect of motif spacing on nucleosome occupancy. The plot shows the occupancy on all the sequences in our library that contain the same two motifs that are 25 bp versus 20 bp apart.

(C) Nucleosome occupancy in the presence of 1–3 copies of Tbf1 or Rsc3 consensus (red) and mutant (blue) motifs. The error bars are the SE from three biological replicates (same as in E).

(D) The effect of motif copy number on the local nucleosome occupancy (±SE) in the native genome. The motifs were divided into "single" or "cluster" (single or multiple sites within 147 bp). Unlike Abf1 and Reb1, Tbf1 and Rsc3 require clustered motifs to cause significant nucleosome displacement.

(E) Deletion test of two group 2 NDFs, Bas1 and Ume6. MNase-qPCR assays were performed on sequences containing three Bas1 or Ume6 motifs (arrows) in the presence or absence of the corresponding TF.

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 binding sites of Rap1 (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

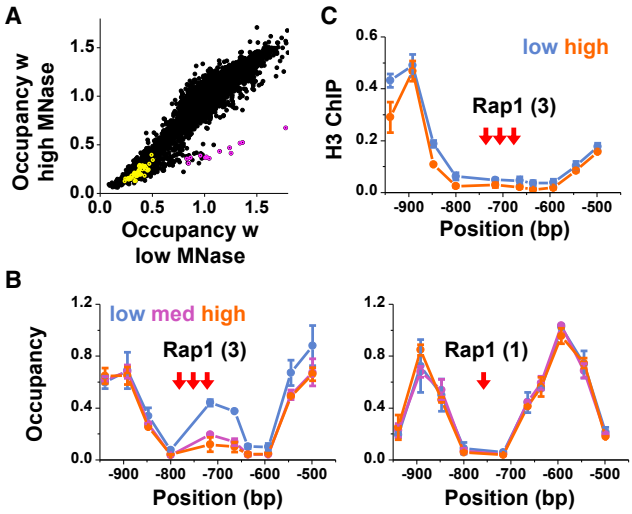


Figure 5. Multiple Rap1 Sites Generate Nucleosome-Sized Particles that Do Not Contain Histones

(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.

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 sites (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 *cbf1*$^{-}$ 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

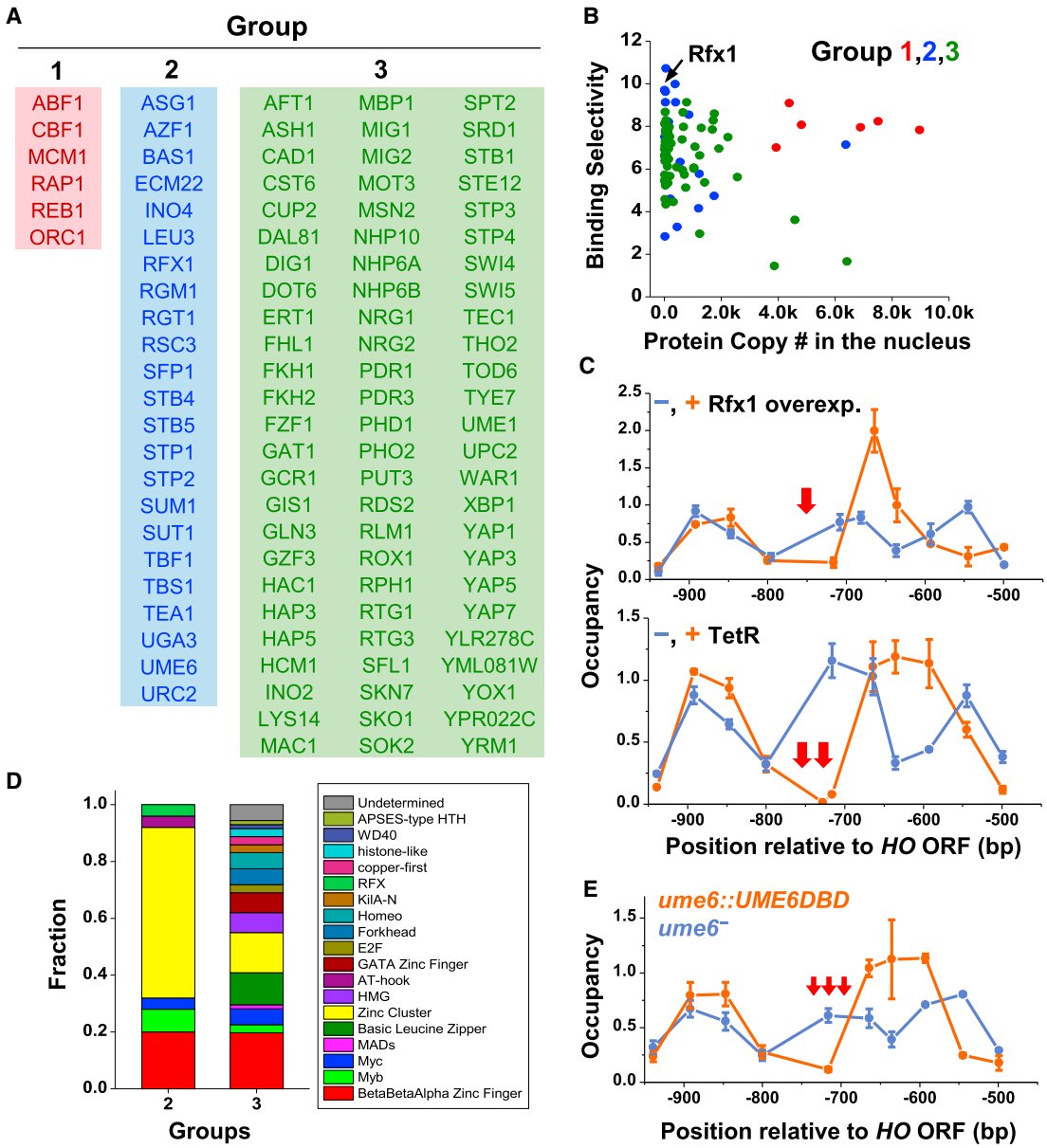


Figure 6. Differences among Three Groups of TFs

(A) List of three groups of TFs. Groups 1 and 2: NDFs that can cause nucleosome displacement with single/multiple motifs; group 3: TFs that are available in the nucleus but have no NDF activities.

(B) Binding affinity versus concentration of the factors in the three groups (different colors). Group 1 factors (red) all have high concentration and binding affinity. The circled dot (pink) in group 2 factor represents Rfx1.

(C) The effect of Rfx1 overexpression (upper panel) and TetR expression (lower panel) on nucleosome occupancy. The DNA templates contain a single Rfx1 site or two TetR sites (arrows) in nucleosome −4, respectively.

(D) Structural analysis of the DNA binding domains (DBDs) of the group 2 and 3 factors. Group 2 is enriched with TFs containing zinc cluster DBDs (yellow).

(E) Ume6 DBD has NDF activity. We deleted a large fraction of the *UME6* gene except its DBD. The DNA templates contains three Ume6 binding sites (same as in Figure 4E) (arrows).

The error bars in (C) and (E) 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

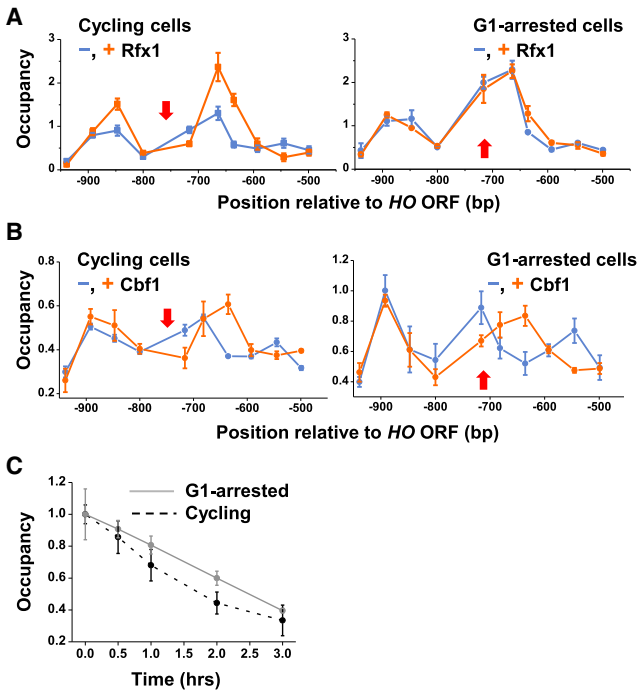


Figure 7. Dependence of Nucleosome Displacement on the Cell Cycle

(A) Nucleosome occupancy with or without Rfx1 overexpression in cycling (left) or G1-arrested (right) cells. The template contains one Rfx1 binding site (red arrow).

(B) Nucleosome occupancy with or without Cbf1 induction in cycling (left) or G1-arrested (right) cells. The template contains one Cbf1 binding site (red arrow) and the "+ Cbf1" curve (red) shows the data with 1 hr Cbf1 induction.

(C) Nucleosome displacement rate by Cbf1 in G1-arrested (gray solid line) versus cycling (black dashed line) cells.

All error bars in this figure are the SE from three biological replicates.

another layer of regulation for NDF function in the native genome. Such complication is avoided in our assay by having the sequence library inserted into the same genomic locus. The drawback of the method is that it requires the prior knowledge of TF motifs, which may not exist in some species. Incorrect or ambiguous motif assignment can cause error in the data interpretation. To confirm the nucleosome-displacing activities of newly identified NDFs, deletion experiment as in Figure 4E needs to be done, which can be time consuming.

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 nucleosome: 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 rebind (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 NDF 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 Hörz, 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.

STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
  - Plasmids and yeast strains
  - Library design and construction
  - Measurement of nucleosome occupancy
  - Amplicon sequencing
  - Sequencing data analysis
  - Further selection of TF motifs
  - Low throughput test on individual TF and template
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Model and Bioinformatics

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and seven tables and can be found with this article online at https://doi.org/10.1016/j.molcel.2018.06.017.

ACKNOWLEDGMENTS

The authors are grateful to Dr. B. Franklin Pugh, Dr. Shaun Mahony, Dr. William K.M. Lai, Nina P. Farrell, Dr. Bongsoo Park, and Kylie Bocklund for help with the amplicon sequencing. We also thank Dr. David Stillman, Dr. Hiten Madhani, Dr. David Clark, and Dr. David MacAlpine for sharing plasmids, strains, and protocols. We acknowledge all members in the Bai lab for insightful comments on the manuscript. We thank the members of the Center of Eukaryotic Gene Regulation at PSU for discussions and technical support. This work is supported by the National Institutes of Health (R01 GM121858).

AUTHOR CONTRIBUTIONS

L.B. and C.Y. designed the experiments; C.Y. and H.C. performed the experiments; C.Y. and L.B. performed the data analysis and bioinformatics analysis; and L.B. and C.Y. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 29, 2017
Revised: May 4, 2018
Accepted: June 7, 2018
Published: July 12, 2018

REFERENCES

Adams, C.C., and Workman, J.L. (1995). Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. Mol. Cell. Biol. *15*, 1405–1421.

Anderson, S.F., Steber, C.M., Esposito, R.E., and Coleman, J.E. (1995). UME6, a negative regulator of meiosis in *Saccharomyces cerevisiae*, contains a C-terminal Zn2Cys6 binuclear cluster that binds the URS1 DNA sequence in a zinc-dependent manner. Protein Sci. *4*, 1832–1843.

Angus-Hill, M.L., Schlichter, A., Roberts, D., Erdjument-Bromage, H., Tempst, P., and Cairns, B.R. (2001). A Rsc3/Rsc30 zinc cluster dimer reveals novel roles for the chromatin remodeler RSC in gene expression and cell cycle control. Mol. Cell *7*, 741–751.

Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L., et al. (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Mol. Cell *32*, 878–887.

Bai, L., Charvin, G., Siggia, E.D., and Cross, F.R. (2010). Nucleosome-depleted regions in cell-cycle-regulated promoters ensure reliable gene expression in every cell cycle. Dev. Cell *18*, 544–555.

Bai, L., Ondracka, A., and Cross, F.R. (2011). Multiple sequence-specific factors generate the nucleosome-depleted region on CLN2 promoter. Mol. Cell *42*, 465–476.

Bi, X., and Broach, J.R. (2001). Chromosomal boundaries in *S. cerevisiae*. Curr. Opin. Genet. Dev. *11*, 199–204.

Boeger, H., Griesenbeck, J., Strattan, J.S., and Kornberg, R.D. (2003). Nucleosomes unfold completely at a transcriptionally active promoter. Mol. Cell *11*, 1587–1598.

Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E.V., Geistlinger, T.R., et al. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell *122*, 33–43.

Chereji, R.V., Ocampo, J., and Clark, D.J. (2017). MNase-sensitive complexes in yeast: nucleosomes and non-histone barriers. Mol. Cell *65*, 565–577.e3.

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K.S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol. Cell *9*, 279–289.

Dechassa, M.L., Sabri, A., Pondugula, S., Kassabov, S.R., Chatterjee, N., Kladde, M.P., and Bartholomew, B. (2010). SWI/SNF has intrinsic nucleosome disassembly activity that is dependent on adjacent nucleosomes. Mol. Cell *38*, 590–602.

Donaghey, J., Thakurela, S., Charlton, J., Chen, J.S., Smith, Z.D., Gu, H., Pop, R., Clement, K., Stamenova, E.K., Karnik, R., et al. (2018). Genetic

determinants and epigenetic effects of pioneer-factor occupancy. Nat. Genet. *50*, 250–258.

Eaton, M.L., Galani, K., Kang, S., Bell, S.P., and MacAlpine, D.M. (2010). Conserved nucleosome positioning defines replication origins. Genes Dev. *24*, 748–753.

Fedor, M.J., Lue, N.F., and Kornberg, R.D. (1988). Statistical positioning of nucleosomes by specific protein-binding to an upstream activating sequence in yeast. J. Mol. Biol. *204*, 109–127.

Garcia, J.F., Dumesic, P.A., Hartley, P.D., El-Samad, H., and Madhani, H.D. (2010). Combinatorial, site-specific requirement for heterochromatic silencing factors in the elimination of nucleosome-free regions. Genes Dev. *24*, 1758–1771.

Gertz, J., Siggia, E.D., and Cohen, B.A. (2009). Analysis of combinatorial cis-regulation in synthetic and genomic promoters. Nature *457*, 215–218.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737–741.

Goldmark, J.P., Fazzio, T.G., Estep, P.W., Church, G.M., and Tsukiyama, T. (2000). The Isw2 chromatin remodeling complex represses early meiotic genes upon recruitment by Ume6p. Cell *103*, 423–433.

Hanna, J., Markoulaki, S., Schorderet, P., Carey, B.W., Beard, C., Wernig, M., Creyghton, M.P., Steine, E.J., Cassady, J.P., Foreman, R., et al. (2008). Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. Cell *133*, 250–264.

Hartley, P.D., and Madhani, H.D. (2009). Mechanisms that specify promoter nucleosome location and identity. Cell *137*, 445–458.

Hsu, H.T., Chen, H.M., Yang, Z., Wang, J., Lee, N.K., Burger, A., Zaret, K., Liu, T., Levine, E., and Mango, S.E. (2015). TRANSCRIPTION. Recruitment of RNA polymerase II by the pioneer transcription factor PHA-4. Science *348*, 1372–1376.

Hurtado, A., Holmes, K.A., Ross-Innes, C.S., Schmidt, D., and Carroll, J.S. (2011). FOXA1 is a key determinant of estrogen receptor function and endocrine response. Nat. Genet. *43*, 27–33.

Iwafuchi-Doi, M., and Zaret, K.S. (2014). Pioneer transcription factors in cell reprogramming. Genes Dev. *28*, 2679–2692.

Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J.A., Mahony, S., Pugh, B.F., Lee, D., Kaestner, K.H., and Zaret, K.S. (2016). The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. Mol. Cell *62*, 79–91.

Kadosh, D., and Struhl, K. (1997). Repression by Ume6 involves recruitment of a complex containing Sin3 corepressor and Rpd3 histone deacetylase to target promoters. Cell *89*, 365–371.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., and Segal, E. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362–366.

Kubik, S., Bruzzone, M.J., Jacquet, P., Falcone, J.L., Rougemont, J., and Shore, D. (2015). Nucleosome stability distinguishes two different promoter types at all protein-coding genes in yeast. Mol. Cell *60*, 422–434.

Lee, C.S., Friedman, J.R., Fulmer, J.T., and Kaestner, K.H. (2005). The initiation of liver development is dependent on Foxa transcription factors. Nature *435*, 944–947.

Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. Nat. Genet. *39*, 1235–1244.

Levo, M., Avnit-Sagi, T., Lotan-Pompan, M., Kalma, Y., Weinberger, A., Yakhini, Z., and Segal, E. (2017). Systematic Investigation of Transcription Factor Activity in the Context of Chromatin Using Massively Parallel Binding and Expression Assays. Mol. Cell *65*, 604–617.e6.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Lin, H.Y., Lin, S.E., Chien, S.F., and Chern, M.K. (2011). Electroporation for three commonly used yeast strains for two-hybrid screening experiments. Anal. Biochem. *416*, 117–119.

Lupien, M., Eeckhoute, J., Meyer, C.A., Wang, Q., Zhang, Y., Li, W., Carroll, J.S., Liu, X.S., and Brown, M. (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell *132*, 958–970.

Maerkl, S.J., and Quake, S.R. (2007). A systems approach to measuring the binding energy landscapes of transcription factors. Science *315*, 233–237.

Magnani, L., Ballantyne, E.B., Zhang, X., and Lupien, M. (2011). PBX1 genomic pioneer function drives ERα signaling underlying progression in breast cancer. PLoS Genet. *7*, e1002368.

Magnani, L., Patten, D.K., Nguyen, V.T., Hong, S.P., Steel, J.H., Patel, N., Lombardo, Y., Faronato, M., Gomes, A.R., Woodley, L., et al. (2015). The pioneer factor PBX1 is a novel driver of metastatic progression in ERα-positive breast cancer. Oncotarget *6*, 21878–21891.

Malik, S., Jiang, S., Garee, J.P., Verdin, E., Lee, A.V., O'Malley, B.W., Zhang, M., Belaguli, N.S., and Oesterreich, S. (2010). Histone deacetylase 7 and FoxA1 in estrogen-mediated repression of RPRM. Mol. Cell. Biol. *30*, 399–412.

Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., and Pugh, B.F. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res. *18*, 1073–1083.

Mirny, L.A. (2010). Nucleosome-mediated cooperativity between transcription factors. Proc. Natl. Acad. Sci. USA *107*, 22534–22539.

Nakshatri, H., and Badve, S. (2007). FOXA1 as a therapeutic target for breast cancer. Expert Opin. Ther. Targets *11*, 507–514.

Ozonov, E.A., and van Nimwegen, E. (2013). Nucleosome free regions in yeast promoters result from competitive binding of transcription factors that interact with chromatin modifiers. PLoS Comput. Biol. *9*, e1003181.

Pataskar, A., Jung, J., Smialowski, P., Noack, F., Calegari, F., Straub, T., and Tiwari, V.K. (2016). NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. EMBO J. *35*, 24–45.

Ramachandran, S., and Henikoff, S. (2016). Transcriptional Regulators compete with nucleosomes post-replication. Cell *165*, 580–592.

Rossetti, L., Cacchione, S., De Menna, A., Chapman, L., Rhodes, D., and Savino, M. (2001). Specific interactions of the telomeric protein Rap1p with nucleosomal binding sites. J. Mol. Biol. *306*, 903–913.

Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., Keren, L., Yakhini, Z., Weinberger, A., and Segal, E. (2012). Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nat. Biotechnol. *30*, 521–530.

Smith, R.P., Taher, L., Patwardhan, R.P., Kim, M.J., Inoue, F., Shendure, J., Ovcharenko, I., and Ahituv, N. (2013). Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. Nat. Genet. *45*, 1021–1028.

Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell *161*, 555–568.

Spivak, A.T., and Stormo, G.D. (2012). ScerTF: a comprehensive database of benchmarked position weight matrices for Saccharomyces species. Nucleic Acids Res. *40*, D162–D168.

Stockdale, C., Flaus, A., Ferreira, H., and Owen-Hughes, T. (2006). Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes. J. Biol. Chem. *281*, 16279–16288.

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nat. Struct. Mol. Biol. *20*, 267–273.

Sun, Y., Sriramajayam, K., Luo, D., and Liao, D.J. (2012). A quick, cost-free method of purification of DNA fragments from agarose gel. J. Cancer *3*, 93–95.

Svaren, J., and Hörz, W. (1997). Transcription factors vs nucleosomes: regulation of the PHO5 promoter in yeast. Trends Biochem. Sci. *22*, 93–97.

Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., et al. (2017). The STRING

database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. *45* (D1), D362–D368.

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

Takahata, S., Yu, Y., and Stillman, D.J. (2009). FACT and Asf1 regulate nucleosome dynamics and coactivator binding at the HO promoter. Mol. Cell *34*, 405–415.

Taneda, T., and Kikuchi, A. (2004). Genetic analysis of RSC58, which encodes a component of a yeast chromatin remodeling complex, and interacts with the transcription factor Swi6. Mol. Genet. Genomics *271*, 479–489.

Tsankov, A., Yanagisawa, Y., Rhind, N., Regev, A., and Rando, O.J. (2011). Evolutionary divergence of intrinsic and trans-regulated nucleosome positioning sequences reveals plastic rules for chromatin organization. Genome Res. *21*, 1851–1862.

van Bakel, H., Tsui, K., Gebbia, M., Mnaimneh, S., Hughes, T.R., and Nislow, C. (2013). A compendium of nucleosome and transcript profiles reveals determinants of chromatin architecture and transcription. PLoS Genet. *9*, e1003479.

Wang, Q., Li, W., Liu, X.S., Carroll, J.S., Jänne, O.A., Keeton, E.K., Chinnaiyan, A.M., Pienta, K.J., and Brown, M. (2007). A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. Mol. Cell *27*, 380–392.

Wapinski, I., Pfeffer, A., Friedman, N., and Regev, A. (2007). Natural history and evolutionary principles of gene duplication in fungi. Nature *449*, 54–61.

Yan, C., Zhang, D., Raygoza Garay, J.A., Mwangi, M.M., and Bai, L. (2015). Decoupling of divergent gene regulation by sequence-specific DNA binding factors. Nucleic Acids Res. *43*, 7292–7305.

Yarragudi, A., Miyake, T., Li, R., and Morse, R.H. (2004). Comparison of ABF1 and RAP1 in chromatin opening and transactivator potentiation in the budding yeast *Saccharomyces cerevisiae*. Mol. Cell. Biol. *24*, 9152–9164.

Yen, K., Vinayachandran, V., Batta, K., Koerber, R.T., and Pugh, B.F. (2012). Genome-wide nucleosome specificity and directionality of chromatin remodelers. Cell *149*, 1461–1473.

Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-scale identification of nucleosome positions in *S. cerevisiae*. Science *309*, 626–630.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241.

Zhang, Y., Moqtaderi, Z., Rattner, B.P., Euskirchen, G., Snyder, M., Kadonaga, J.T., Liu, X.S., and Struhl, K. (2009). Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. Nat. Struct. Mol. Biol. *16*, 847–852.

Zhang, Z., Wippo, C.J., Wal, M., Ward, E., Korber, P., and Pugh, B.F. (2011). A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. Science *332*, 977–980.

Zhang, Q., Yoon, Y., Yu, Y., Parnell, E.J., Garay, J.A., Mwangi, M.M., Cross, F.R., Stillman, D.J., and Bai, L. (2013). Stochastic expression and epigenetic memory at the yeast HO promoter. Proc. Natl. Acad. Sci. USA *110*, 14012–14017.

Zhou, X., and O'Shea, E.K. (2011). Integrated approaches reveal determinants of genome-wide binding and function of the transcription factor Pho4. Mol. Cell *42*, 826–836.

Zhu, X., and Keeney, S. (2015). High-resolution global analysis of the influences of Bas1 and Ino4 transcription factors on meiotic DNA break distributions in *Saccharomyces cerevisiae*. Genetics *201*, 525–542.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Histone H3 C-terminal Antibody	Epicypher	Cat# 13-0001
Bacterial and Virus Strains		
NEB 5-alpha Competent *E. coli* (High Efficiency)	New England Biolabs	Cat# C2987
NEB 5-alpha Electrocompetent *E. coli*	New England Biolabs	Cat# C2989
Chemicals, Peptides, and Recombinant Proteins		
nuclease micrococcal from Staphylococcus aureus	Sigma	Cat# N3755
D-(+)-Galactose	Sigma	Cat# G0750
D-(+)-Raffinose pentahydrate	Sigma	Cat# 83400
Protease Inhibitor Cocktail	Sigma	Cat# P8215
Protein A/G PLUS-Agarose	Santa Cruz Biotechnology	Cat# sc-2003
Protainase K	VWR	Cat# 97062-670
Z1005 Zymolyase	United States Biological	Cat# 37340-57-1
Experimental Models: Organisms/Strains		
yLB91	This study (Table S5)	N/A
yCY32-Ume6	This study (Table S5)	N/A
yCY32-Bas1	This study (Table S5)	N/A
yCY32-Rgt1	This study (Table S5)	N/A
yCY32-Uga3	This study (Table S5)	N/A
yCY32-Mbp1	This study (Table S5)	N/A
yCY32-Reb1	This study (Table S5)	N/A
yCY32-Reb1mut	This study (Table S5)	N/A
yCY32-Cbf1	This study (Table S5)	N/A
yCY57-Ume6	This study (Table S5)	N/A
yCY57-Bas1	This study (Table S5)	N/A
yCY57-Rgt1	This study (Table S5)	N/A
yCY57-Uga3	This study (Table S5)	N/A
yCY50	This study (Table S5)	N/A
yCY55-Rfx1	This study (Table S5)	N/A
yCY55-tetR	This study (Table S5)	N/A
yCY62-Rap1	This study (Table S5)	N/A
yCY62-3Rap1	This study (Table S5)	N/A
yCY61-Rfx1	This study (Table S5)	N/A
yCY61-TetR	This study (Table S5)	N/A
yCY63-Rfx1	This study (Table S5)	N/A
yCY65	This study (Table S5)	N/A
yCY66	This study (Table S5)	N/A
yCY67	This study (Table S5)	N/A
Oligonucleotides		
Library Amplification Primers	This study (Table S6)	N/A
Designed oligo library	This study (Table S3)	N/A
Recombinant DNA		
pCY15	This study	N/A
pCY20	This study	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pCY32	This study	N/A
pCY36	This study	N/A
pCY35	This study	N/A
Software and Algorithms		
Burrows-Wheeler Aligner	Li and Durbin, 2009	RRID:SCR_010910
FASTX-Toolkit	Cold Spring Harbor Laboratory	RRID:SCR_005534
MATLAB	Math Works	RRID:SCR_001622

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* −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 (~$1E^{10}$ 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 BbsI digestion at 37°C overnight (Table S6). The 60bp band in the digestion product was gel-purified (Sun et al., 2012) and ligated to BbsI-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 × 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-*. We grew yCY63-Rfx1 in the repressive condition (+methionine) and arrested the cells in G1 with alpha factor for 1.5 hr. We induced Rfx1 overexpression for 2 hr while maintaining alpha factor in the media. Control cells were subjected to the same treatment but without alpha factor. The nucleosome occupancies were measured before and after the Rfx1 overexpression. 3) MNase-histone ChIP: we modified our previous MNase assay to combine it with the immunoprecipitation. Briefly, we crosslinked the culture with 1% formaldehyde for 15min and quenched with 0.125mM glycine for 5min. We performed the MNase assay as described (Bai et al., 2011) except that we terminated the MNase digestion with 0.025mM EDTA and released chromatin by adding 0.5% SDS. The chromatin was then subject to solution exchange for RIPA buffer without SDS. After standard ChIP procedure with anti-H3 (epicypher 13-0001), the resulting DNA was analyzed using stacking qPCR. All ChIP signals were normalized to that of nucleosome −1 on *PHO5* promoter. 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.