Molecular Cell
Short Article

Integrated Approaches Reveal Determinants of Genome-wide Binding and Function of the Transcription Factor Pho4

Xu Zhou[1,2] and Erin K. O'Shea[1,2,3,*]
[1]Howard Hughes Medical Institute
[2]Department of Molecular and Cellular Biology
[3]Department of Chemistry and Chemical Biology
Harvard University Faculty of Arts and Sciences Center for Systems Biology, Northwest Labs, 52 Oxford Street, Cambridge, MA 02138, USA
*Correspondence: erin_oshea@harvard.edu
DOI 10.1016/j.molcel.2011.05.025

SUMMARY

DNA sequences with high affinity for transcription factors occur more frequently in the genome than instances of genes bound or regulated by these factors. It is not clear what factors determine the genome-wide pattern of binding or regulation for a given transcription factor. We used an integrated approach to study how *trans* influences shape the binding and regulatory landscape of Pho4, a budding yeast transcription factor activated in response to phosphate limitation. We find that nucleosomes significantly restrict Pho4 binding. At nucleosome-depleted sites, competition from another transcription factor, Cbf1, determines Pho4 occupancy, raising the threshold for transcriptional activation in phosphate replete conditions and preventing Pho4 activation of genes outside the phosphate regulon during phosphate starvation. Pho4 binding is not sufficient for transcriptional activation—a cooperative interaction between Pho2 and Pho4 specifies genes that are activated. Combining these experimental observations, we are able to globally predict Pho4 binding and its functionality.

INTRODUCTION

Specific transcriptional regulation is essential for precise control of many biological processes, including tissue development and responses to environmental stimuli. Alterations in gene expression are commonly associated with human diseases and disorders (Jimenez-Sanchez et al., 2001; Vaquerizas et al., 2009). A key regulatory step in the control of gene expression is the binding of transcription factors to DNA to activate gene transcription (Hochheimer and Tjian, 2003; Ptashne and Gann, 1997). Both *cis* and *trans* factors can influence transcription factor binding. In *cis*, binding typically requires recognition of a specific DNA sequence (Jacob and Monod, 1961); sequence variation in DNA binding motifs affects transcription factor binding occupancy (Kasowski et al., 2010; Schmidt et al., 2010; Zheng et al., 2010). However, high-affinity binding sites occur more frequently than experimentally detected binding events, even for transcription factors in organisms with a relatively small genome such as *Saccharomyces cerevisiae* (Harbison et al., 2004; MacIsaac et al., 2006). Among vertebrates, binding events display a species-specific pattern despite similarity between the consensus binding motifs of transcription factors (Schmidt et al., 2010). Among human individuals, the majority of transcription factor binding variance cannot be explained by genetic differences in binding sites (Kasowski et al., 2010), suggesting that *trans* factors also influence transcription factor binding. In *trans*, chromatin structure restricts DNA accessibility (Liu et al., 2006; Wasson and Hartemink, 2009), but the effect of nucleosome occupancy on the sequence-specific binding of transcription factors has not yet been systematically explored. Other *trans* factors, such as cooperating and competing factors (Pan et al., 2010; Pierce et al., 2003), have been studied primarily at the level of individual genes for their influence on transcriptional control, but not evaluated on a global scale. Moreover, binding of a transcription factor per se is often not sufficient for transcriptional regulation (Birney et al., 2007; Farnham, 2009; Harbison et al., 2004; MacIsaac et al., 2006). Thus, it remains unclear what determines the genomic locations to which a transcription factor binds and whether this binding is able to influence the transcription of a gene (Farnham, 2009).

To further complicate matters, transcription factors of the same family contain structurally conserved DNA binding domains and usually recognize similar short DNA motifs (Badis et al., 2009; Berger et al., 2008; Wei et al., 2010). However, these factors frequently regulate distinct biological responses (Robinson and Lopes, 2000; Sharrocks, 2001). The conflict between similarity in binding and divergence in regulation raises two questions: How are distinct patterns of transcriptional regulation achieved? How is the regulation by a given transcription factor influenced by others that recognize very similar DNA motifs? To answer these questions, it is necessary to systematically interrogate, on a genome-wide scale, the factors that may contribute to the specificity of transcription factor binding and regulation.

We used gene regulation by the phosphate-responsive (*PHO*) signaling pathway as a model system to explore the determinants for transcription factor binding and function. Pho4, a basic-helix-loop-helix (bHLH) transcriptional activator in

S. cerevisiae, is regulated in response to environmental inorganic phosphate (Pi) availability and activates a transcriptional program together with the transcription factor Pho2 when cells are limited for Pi (Ogawa et al., 2000). In Pi rich conditions, Pho4 is phosphorylated and transported from the nucleus into the cytoplasm; in phosphate-limited conditions, Pho4 is dephosphorylated and transported from the cytoplasm into the nucleus (O'Neill et al., 1996; Schneider et al., 1994). Pho4 binds with high affinity to "CACGTG" motifs in vivo (Harbison et al., 2004) and in vitro (Badis et al., 2008; Maerkl and Quake, 2007; Zhu et al., 2009). In this study, we found that Pho4 is bound to only ~14% of its consensus binding sites, and only a quarter of those binding events result in activation of gene transcription. We demonstrated that this genome-wide binding and regulatory pattern is determined by a combination of chromatin restriction, competitive binding from the factor Cbf1 that recognizes the same motif as Pho4, and cooperation with Pho2. We also discovered that Cbf1 competition raises the threshold of gene activation and helps to define the specificity of *PHO* gene regulation. Combining our experimental observations, we can predict transcription factor binding and function at a whole genome level.

RESULTS

Chromatin Negatively Regulates Pho4 Binding

We applied biotin-tagging chromatin immunoprecipitation (Kolodziej et al., 2009; van Werven and Timmers, 2006) combined with high-throughput sequencing (Bio-ChIP-Seq; see the Experimental Procedures) to identify Pho4 binding events in vivo in phosphate starvation conditions (Figure 1A and Figure S1A available online). Although there are 843 "CACGTG" consensus binding sites (all "NCACGTGN" motifs except "TCACGTGA" are considered to be consensus binding sites; see the Experimental Procedures for more details) in the genome, we observed Pho4 binding to only 115 (~14%) of these sites (Experimental Procedures), implying that factors other than DNA binding specificity influence Pho4 binding in vivo. Since chromatin restricts the access of transcription factors to their potential binding sites (Khorasanizadeh, 2004; Kornberg and Lorch, 1999; Narlikar et al., 2002), a Pho4 "CACGTG" consensus binding site could exist in one of two possible states: an inaccessible state, occluded by nucleosomes; or an accessible state, exposed in a nucleosome-depleted region (nucleosome-free or nucleosome linker region). To determine whether local chromatin structure influences transcription factor binding, we mapped nucleosome occupancy in no-Pi conditions with micrococcal nuclease digestion followed by paired-end deep sequencing (Figure 1B, Figures S1B–S1E, and the Experimental Procedures). As expected, Pho4 is not bound to the binding sites that are inaccessible and most occluded by nucleosomes (see the Experimental Procedures for details) (209 of 216 sites, 97%; Figure 2A and Figure S2A). However, with the same threshold, Pho4 is also not bound to two thirds of the most accessible sites (172 of 248 sites) (Figure 2A and Figure S2A). We conclude that chromatin structure inhibits transcription factor binding, but open chromatin structure is not sufficient for Pho4 binding at its "CACGTG" consensus sites.

Competition from Cbf1 Determines Pho4 Occupancy In Vivo at Nucleosome-Depleted Sites

It is possible that proteins with similar specificity compete with Pho4 for binding to accessible "CACGTG" consensus sites. Cbf1, another member of the bHLH transcription factor family, is present in the nucleus at high concentration (Ghaemmaghami et al., 2003), is not known to interact with Pho4 (Graumann et al., 2004), and binds with high affinity to the same consensus binding motif "CACGTG" in vitro and in vivo as does Pho4 (Harbison et al., 2004; MacIsaac et al., 2006; Maerkl and Quake, 2007; Zhu et al., 2009). To test whether Cbf1 competes with Pho4 for binding, we identified in vivo binding sites for Cbf1 in high- and no-Pi conditions. In no-Pi conditions, 77% (132 of 172) of the accessible consensus "CACGTG" binding sites not bound by Pho4 are occupied by Cbf1 (Figure 2B). Intriguingly, most of the accessible sites bound by Pho4 (72 of 76, 95%) are also bound by Cbf1. Thus, the accessible, high-affinity "CACGTG" sites mainly fall into two classes: those in which Cbf1 competes with Pho4 most effectively, resulting in detectable binding of Cbf1 but not Pho4, and those in which Cbf1 competes less effectively, resulting in significant occupancy of both Pho4 and Cbf1. In high-Pi conditions, Cbf1 is bound to both of these classes of sites (119 of 132 and 67 of 72, respectively; Figure S2B).

Pho4 and Cbf1 have different preferences for bases flanking the "CACGTG" consensus binding site in vitro (Maerkl and Quake, 2007); these sequence features might explain differences in Cbf1 and Pho4 occupancy in vivo. In accord with the observed in vitro sequence preferences, we find that accessible sites with less Pho4 binding have a single 5′ "T" base flanking the "CACGTG" (Figure 2C), suggesting that Cbf1 can compete most effectively at these sites (Figure 2D).

Competitive Binding of Cbf1 Influences the Activation Threshold and Specificity of the *PHO* Regulon

What is the physiological role of the interplay between Pho4 and Cbf1 at consensus sites? One idea is that Cbf1 is required in high-Pi conditions to keep nucleosomes properly positioned, and therefore keep consensus binding sites nucleosome free (Figure S3A; "Cbf1 priming model"). Although there is precedent for Cbf1 positioning nucleosomes in the promoters of some sulfur metabolism genes (Kent et al., 2004; Kent et al., 1994), when we analyzed the *cbf1Δ* strain in high-Pi conditions we observed no change in nucleosome occupancy at Pho4 binding sites in Pho4-regulated genes ($r = 0.943$ between *cbf1Δ* strain and wild-type, $r = 0.953$ between wild-type replicates; Figure S3B); in contrast, in the *cbf1Δ* strain nucleosome occupancy increases and nucleosome position shifts at binding sites in the regulatory regions of sulfur metabolism genes (Lee et al., 2010) (Figure S3C).

An alternative model is that competition from Cbf1 prevents spurious activation of phosphate-responsive genes in high-Pi conditions, and of other "CACGTG"-containing genes during phosphate starvation (Figure S3D; "Cbf1 blocking model"). In the absence of Cbf1, most Pho4-regulated genes (see definition in Figure 4A) showed significantly increased expression in high-Pi medium (20 of 28, $p \leq 0.05$; Figure 3A, column 2)—conditions in which Pho4 is less active and localized primarily to the cytoplasm (Komeili and O'Shea, 1999). This aberrant expression is Pho4 dependent (compare Figure 3A, columns 2 and 3) and is

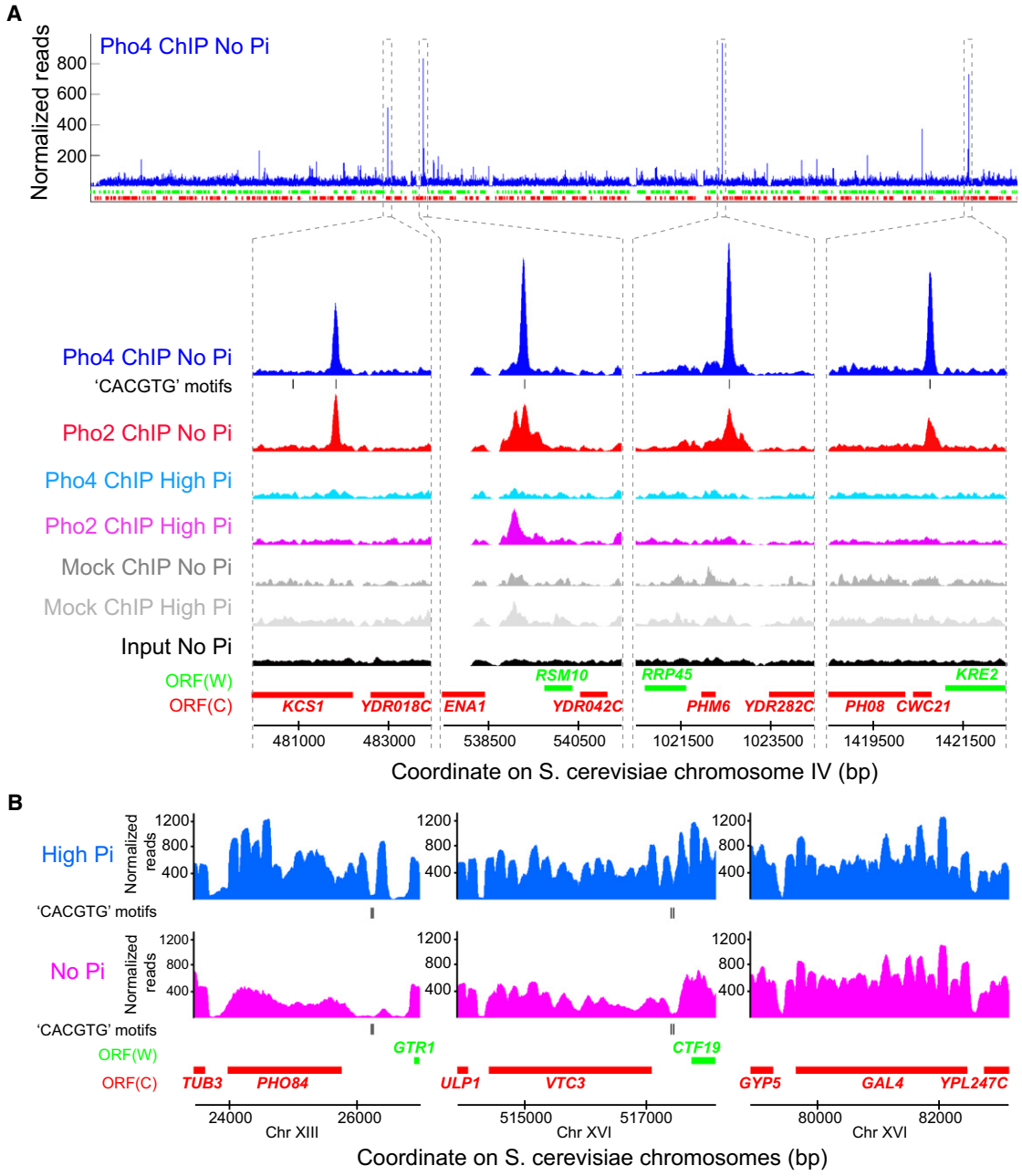

Figure 1. ChIP-Seq Analysis and Examples of Nucleosome Positioning in High- and No-Pi Conditions

(A) ChIP-seq analysis of Pho2 and Pho4 binding and a mock IP (wild-type strain with integrated *E.coli birA* gene) in high- and no-inorganic-phosphate (Pi) conditions, and genomic input in no-Pi conditions. All ChIP results were normalized to Pho4 ChIP at no-Pi conditions (Supplemental Experimental Procedures).

(B) Nucleosome occupancy map for *PHO84* and *VTC3* (both derepressed in no-Pi conditions), *GAL4* (not responsive to Pi concentration), in high- and no-Pi conditions. The nucleosome maps in both conditions were normalized to have the same number of total reads.

See also Figure S1.

not the indirect result of an upstream signaling defect caused by deletion of Cbf1 (Figure S3E). Moreover, deletion of Rtg3 and Tye7, two other members of the bHLH family that bind the "CACGTG" consensus site (Zhu et al., 2009), did not result in spurious activation (Figure 3A, column 4). Since deletion of Cbf1 causes a growth defect and impaired activation of the *PHO* pathway in no-Pi conditions (Figure S3E), we evaluated the consequences of Cbf1 competition when the *PHO* pathway is fully activated using a strain lacking the cyclin Pho80 grown in high-Pi medium (O'Neill et al., 1996). In the *cbf1Δ pho80Δ* strain, we observed Pho4 binding to sites bound by Cbf1 in the wild-type strain (Figure 3D), particularly at the "T-CACGTG"

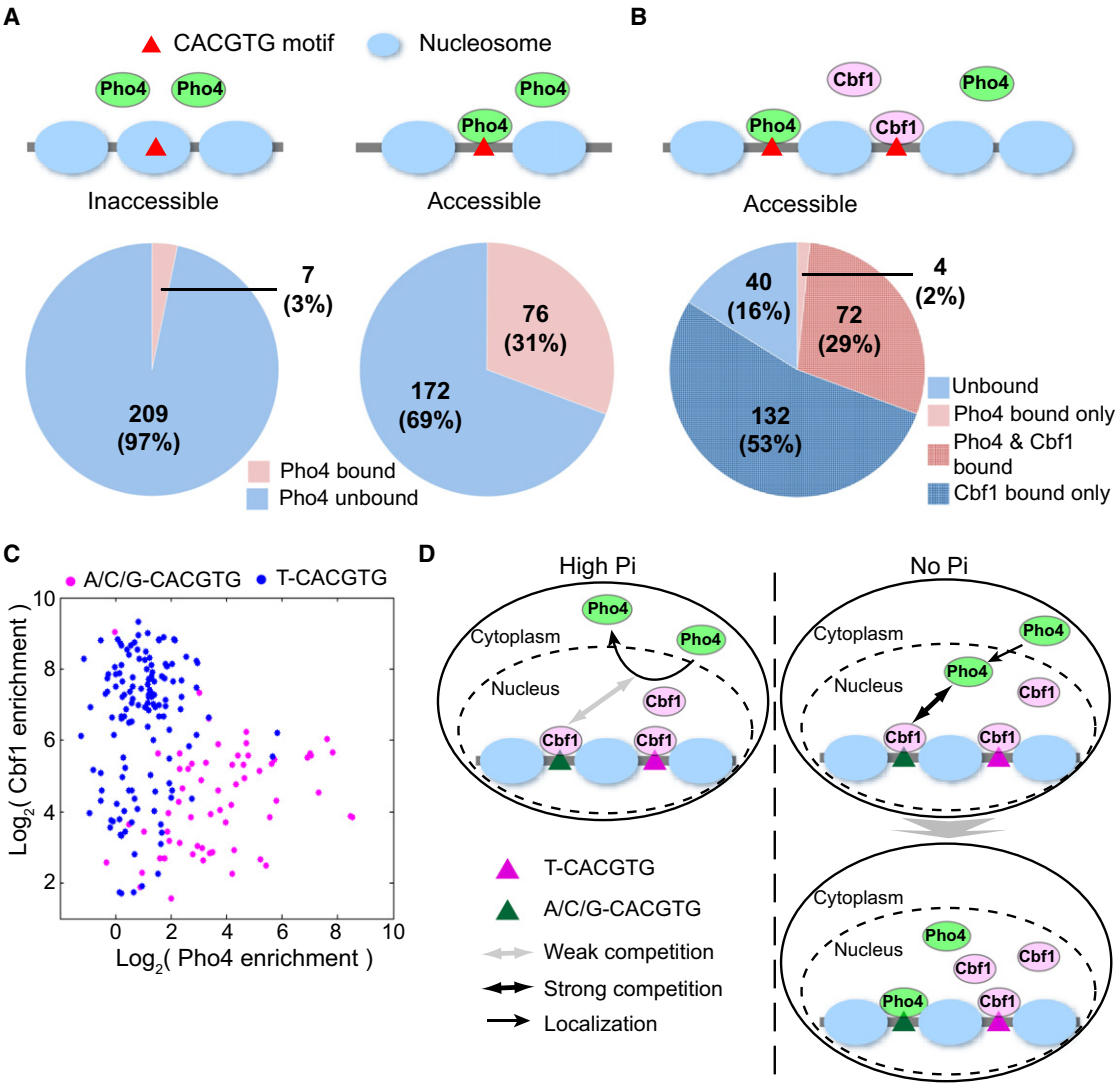

Figure 2. Determinants of Pho4 Binding at Its Consensus Binding Motifs

(A) Pie charts showing Pho4 binding at inaccessible (the quartile most occluded by nucleosomes) and accessible (the quartile least occupied by nucleosomes) consensus binding sites in no-Pi conditions.

(B) Pie chart showing Pho4 and Cbf1 binding at accessible consensus binding sites in no-Pi conditions.

(C) Scatter plot displaying Pho4 and Cbf1 binding enrichment in no-Pi conditions for accessible sites that are bound by at least one of the two transcription factors.

(D) Schematic depicting a model for determinants of Pho4 binding to its consensus motifs. In high-Pi conditions (left), Pho4 is imported into the nucleus and then actively exported; thus, the Pho4 nuclear concentration is low (Komeili and O'Shea, 1999; O'Neill et al., 1996). Most of the accessible consensus binding sites are occupied by Cbf1, which resides in the nucleus constitutively (Huh et al., 2003). In no-Pi conditions (upper right), Pho4 is no longer exported from the nucleus and the nuclear concentration of Pho4 increases, allowing Pho4 to compete effectively for binding at sites that are weakly bound by Cbf1 (lower right); Cbf1 is bound to the consensus sites with a 5′ flanking "T" with high enough affinity to prevent Pho4 binding.

See also Figure S2 and Table S2.

motifs (Figure 3C and Figure S3F). Indeed, Pho4 is bound to and activates transcription of 13 genes that were bound by Cbf1 and not regulated by Pho4 in the wild-type strain (Figure 3B). Our observations support two roles for Pho4-Cbf1 competition: in high-Pi conditions, Cbf1 prevents spurious activation of the *PHO* genes induced by a low level of nuclear Pho4, ensuring that phosphate-responsive genes are turned off when Pi is available; in no-Pi conditions, Cbf1 prevents Pho4 from inappropriately activating genes containing a "CACGTG" motif that are not part of the phosphate regulon, ensuring that only genes needed for the response to phosphate limitation are turned on.

Cooperative Binding between Pho2 and Pho4 Determines the Functionality of Pho4 Binding Events

Pho4 binds to the "CACGTG" motif in the promoters of over 80 genes, but only ~10–20 genes are regulated by Pho4 in

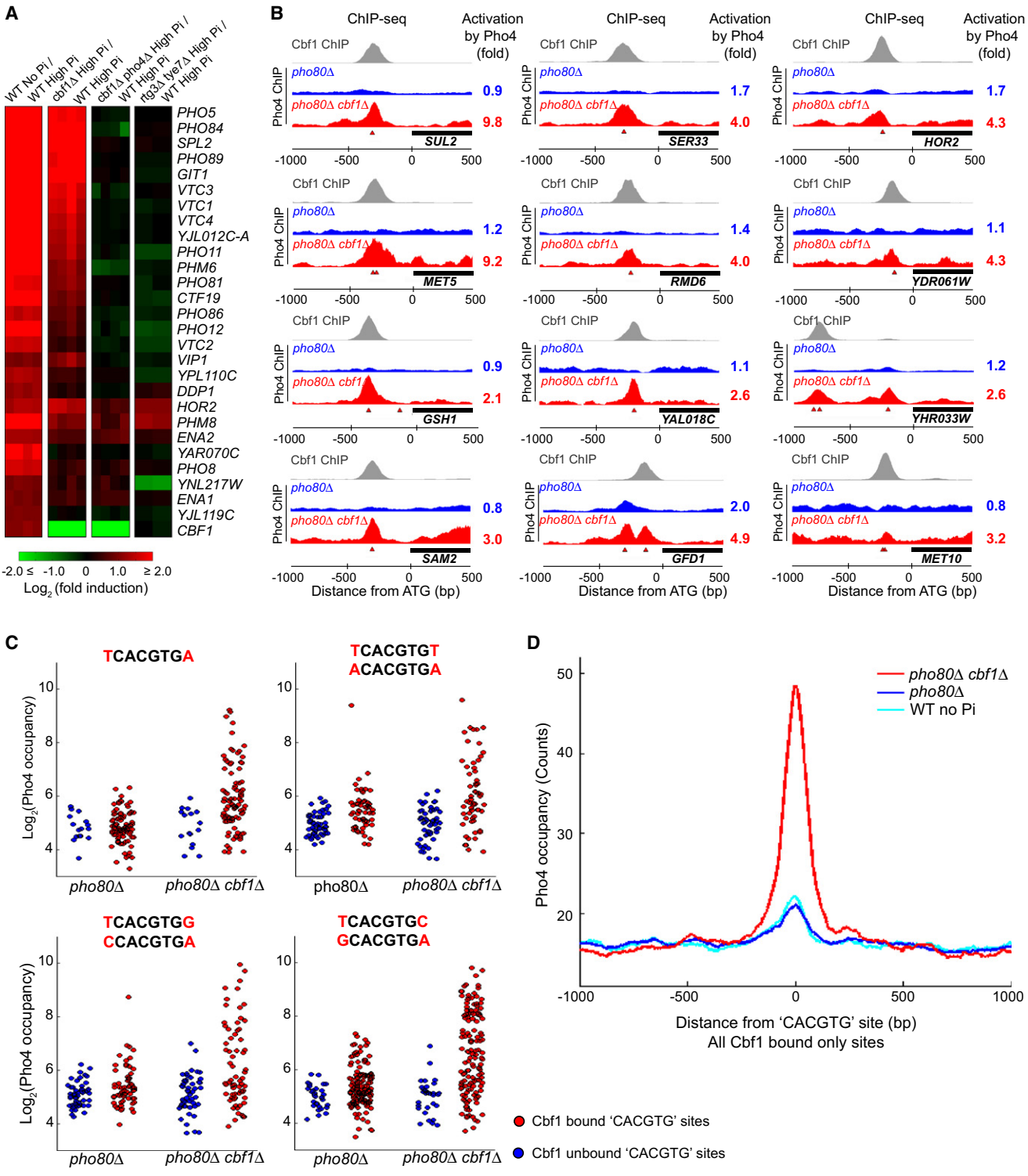

Figure 3. Cbf1 Prevents Pho4 Spurious Transcriptional Sctivation in High-Pi Conditions and Inappropriate Gene Activation in No-Pi Conditions

(A) Heat map showing the fold induction (log_2 scale) of Pho4-regulated genes for wild-type in no-Pi conditions (column 1) and *cbf1Δ*, *cbf1Δ pho4Δ*, and *rtg3Δ tye7Δ* strains in high-Pi conditions (columns 2–4), as measured by microarray analysis. Subcolumns indicate biological replicates.

(B) Differential binding and activation by Pho4 in the presence and absence of Cbf1. Tracks show ChIP-seq results of Cbf1 binding in no-Pi conditions (gray), Pho4 binding in a *pho80Δ* strain (constitutively nuclear Pho4, blue) and in a *cbf1Δ pho80Δ* (red) strain. Gene activation by Pho4 in *pho80Δ* (blue) and *cbf1Δ pho80Δ*

response to Pi starvation (Ogawa et al., 2000; Springer et al., 2003). Moreover, these regulated genes have different dependence on Pho4 and on Pho2, a homeodomain factor that interacts with Pho4 and regulates the phosphate starvation response (Vogel et al., 1989). We used epistasis expression analysis (mutant cycle analysis) (Capaldi et al., 2008) to dissect and quantify the contribution of the regulatory interactions between Pho2 and Pho4 to transcriptional activation (Figure S4) in terms of three ''expression components'': the contribution of Pho2 acting alone (Pho2), the contribution of Pho4 acting alone (Pho4), and the contribution of Pho2 and Pho4 acting together (which we refer to as a ''cooperative'' component). To determine the values of these expression components, we directly compared gene expression between all possible pairs of wild-type, *pho2Δ*, *pho4Δ*, and *pho2Δ pho4Δ* strains in no- and high-Pi conditions, in which the measured gene expression difference between each pair of strains consists of a linear sum of the expression components relevant for that pairwise comparison (Figure S4). We calculate the expression components for each gene through regression methods with equations describing the expression components measured in each microarray (Supplemental Experimental Procedures). We observe only three regulatory interactions operating at phosphate-responsive genes: gene activation by Pho4 acting alone (Pho4 only), genes activated by the interaction between Pho2 and Pho4 (Cooperative, Co), and mixed regulation (genes partially activated by Pho4 alone and partially dependent on the interaction between Pho2 and Pho4, Pho4 + Co) (Figure 4A). Notably, almost all Pho4 regulated genes contain a cooperative component, suggesting that the interaction between Pho2 and Pho4 is necessary for gene activation.

Since Pho2 and Pho4 interact and bind cooperatively to the *PHO5* promoter (Barbaric et al., 1998; Barbarić et al., 1996; Vogel et al., 1989), we hypothesized that this cooperative binding might be correlated with the functionality of Pho4 binding events—the ability to trigger gene activation. We observed a strong correlation between Pho2 and Pho4 binding events in no-Pi conditions ($r = 0.926$, $p < 10^{-49}$; Figure S5A). Only some of these coincident binding events appear to be instances of cooperative binding (see the Experimental Procedures), where both Pho2 and Pho4 occupancy increases in response to phosphate starvation—it is these cooperative binding events that correlate with gene activation (''regulated'' class in Figure 4B, $p = 8.1 \times 10^{-16}$, Fisher's exact test). Recruitment of Pho4 to cooperative, regulated sites is entirely dependent on Pho2, whereas Pho4 binding to the noncooperative, nonregulated sites is largely unaffected by deletion of Pho2 (Figure 4C). The reduction of Pho4 occupancy in *pho2Δ* strains in no-Pi conditions is correlated with the transcriptional activation contributed by the cooperative interaction between Pho2 and Pho4 (Co component) for Pho4-regulated genes ($R = 0.63$, $p = 0.0016$). For the cooperative Pho4 binding events, the Pho2 and Pho4 ChIP signals are overlapping (Figure 4D), and the predicted Pho2 binding sites are enriched at a distance of 15 bp from the consensus Pho4 binding sites (Figure 4E). In contrast, when only one of the factors is significantly recruited in response to Pi limitation, the gene is not activated (''nonregulated'' class in Figure 4B, $p = 0.99$, binomial test), there is no juxtaposition of Pho2 and Pho4 ChIP signals (Figure 4D), and predicted Pho2 binding sites do not exhibit a consistent spatial relationship with Pho4 binding sites (Figure 4E). Thus, the spatial organization of Pho2 and Pho4 binding motifs may promote cooperative binding of Pho2 and Pho4 and the ability to activate transcription.

Prediction of Pho4 Binding and Function

Prediction of transcription factor binding and functional targets has been a challenging task. Here, we provide an integrated mechanistic view of the determinants that influence Pho4 binding and regulation, taking into account all ''CACGTG'' sites independent of evolutionary conservation, clustering of motifs, and relative positioning in the promoter. Incorporating the influence of *trans* effects into an equilibrium binding model (see the Experimental Procedures), 43 of 50 (86%) (Figure 4F) of the top predicted binding sites are indeed bound by Pho4 (AUC-ROC = 0.87). We find that all *trans* effects contribute significantly to the pattern of Pho4 binding in the genome (Figure S5C). Either nucleosome occupancy or flanking sequences predicts Pho4 binding to high-affinity consensus sites (top panel; AUC-ROC). However, if we consider only the group of top predicted targets (bottom two panels), a prediction based on nucleosome occupancy or Cbf1 competition alone has poor accuracy, whereas the synergy between Cbf1 competition and nucleosome occlusion more accurately predicts binding to this group of target sites (Figure S5C and the Supplemental Experimental Procedures). Of the 115 experimentally determined Pho4 binding sites, fewer than 25% are able to promote activation of gene transcription in no-Pi conditions. Considering the cooperative interaction between Pho2 and Pho4, 23 of 28 (82%) (Figure 4F) binding sites predicted to be functional are actually associated with transcriptional activation (AUC-ROC = 0.992). From this prediction, we identified three nonpromoter binding events that are associated with Pi-dependent antisense transcripts (B. Zid and E.K.O., unpublished data)—two of them have not been previously identified and are potentially linked to phosphate signaling. We find that the presence of a Pho2 binding motif 15 bp from the Pho4 binding site is highly predictive of the functionality of the Pho4 binding events (Figure S5D), indicating that the spatial arrangement of Pho2 and Pho4 binding sites is critical in activating gene transcription.

(red) strains is determined by comparing gene expression in *pho80Δ* and *pho80Δ pho4Δ* strains and *cbf1Δ pho80Δ* and *cbf1Δ pho80Δ pho4Δ* strains, respectively. Red triangles mark the consensus ''CACGTG'' binding sites.

(C) Scatter plots show Pho4 binding occupancy at all ''TCACGTG'' sites, both accessible and those occluded by nucleosomes. Labeled 8-mer sequences indicate DNA motifs with the same binding preference of Pho4.

(D) Average Pho4 binding occupancy at sites that are Cbf1-bound only in the wild-type.

See also Figure S3 and Table S2.

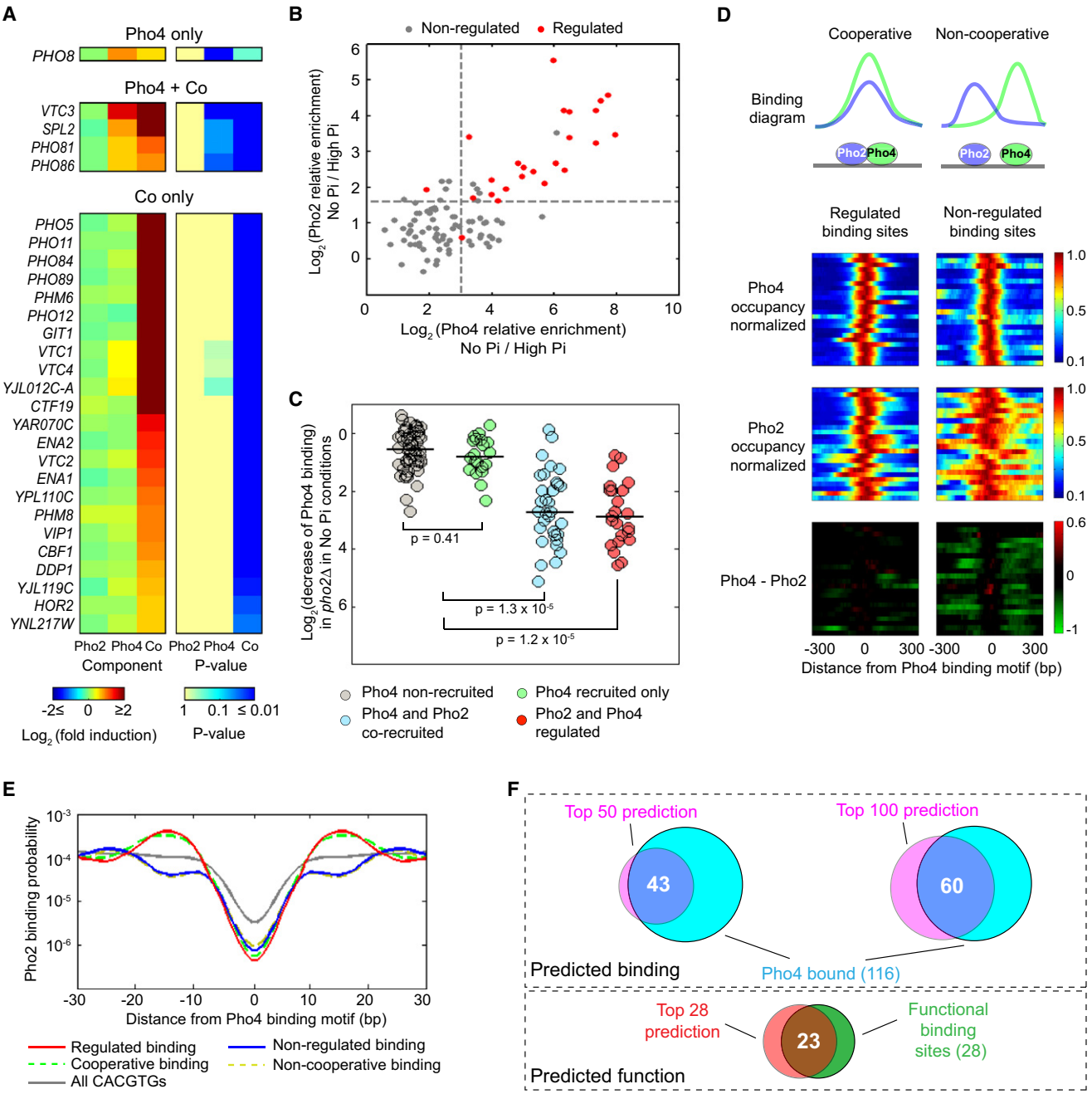

Figure 4. Mutant Cycle Analysis and Correlation between Pho2 and Pho4 Binding and Gene Induction

(A) Heat map showing the best fit of expression components (left columns) and their statistical significance (right columns) for genes that regulated by Pho2 and Pho4. Genes are clustered according to the statistical significance ($p \leq 0.05$) of the Pho2, Pho4, and cooperative component (Co).

(B) Scatter plot showing recruitment ($enrichment_{high}/enrichment_{no}$) of Pho2 and Pho4 after Pi starvation for all Pho4-bound consensus sites within 800 bp of the transcription start site (TSS). "Regulated binding sites" are sites associated with Pho4 regulated genes (Figure 4A). Gray dashed lines indicate the recruitment threshold as mean + 2 standard deviations (SDs) for all consensus motifs excluding the regulated sites.

(C) Plots showing fold decrease in Pho4 occupancy at Pho4-bound consensus binding sites in *pho2Δ* strains. Black lines indicate the median, and p values are calculated with two-sample t tests with unequal variance.

(D) Heat map displaying normalized ChIP occupancy of Pho2 and Pho4 in no-Pi conditions for all sites showing Pho4 recruitment ($\geq$ mean + 2 SDs).

(E) Spatial organization of Pho2 binding motifs predicted based on in vitro binding specificity (Zhu et al., 2009).

(F) Venn diagram showing prediction of Pho4 binding and function derived from a model that incorporates competition and cooperativity.

See also Figures S4 and S5 and Tables S2 and S3.

DISCUSSION

Sequence-specific transcriptional factors recognize short *cis* elements in promoter regions to regulate gene transcription (Farnham, 2009; Jacob and Monod, 1961). However, these elements exist tens of times more frequently than the number of in vivo binding events. How can transcription factors bind to and regulate a specific set of genes when many other genes carry the same binding sites? We used a genome-wide approach to systematically dissect the determinants of transcription factor binding and regulation, and investigated whether these determinants can explain transcriptional specificity. We provide evidence that competition from chromatin, competition from another transcription factor that recognizes Pho4 consensus binding motifs, and cooperativity with a partner protein determine the specificity of Pho4 binding and regulation. We also discovered novel regulatory functions for Cbf1 competition and provide a general view of the specific regulation of the *PHO* regulon.

In high-Pi conditions, Pho4 is phosphorylated and exported to the cytoplasm (O'Neill et al., 1996; Schneider et al., 1994). Competition from Cbf1 increases the threshold of transcriptional activation by Pho4, preventing spurious activation by low levels of nuclear Pho4 (Figure 2D). In no-Pi conditions, phosphorylation of Pho4 is inhibited (O'Neill et al., 1996; Schneider et al., 1994) and the increase in the concentration of unphosphorylated nuclear Pho4 enables it to compete effectively with Cbf1 at sites where Cbf1 is weakly bound—those with certain flanking sequences found in genes within the phosphate regulon; inappropriate binding and activation of "CACGTG"-containing genes outside the phosphate regulon is prevented. The dynamic range of binding and gene activation is increased as a result of the reduction in Pho4 binding due to Cbf1 competition in high-Pi conditions and its condition-dependent cooperative interaction with Pho2 in no-Pi conditions (Pho2 interacts only with unphosphorylated Pho4 during phosphate limitation) (Komeili and O'Shea, 1999). Other stress- and nutrient-response transcriptional programs may employ similar strategies to trigger high levels of induction of specific genes that were tightly repressed in the absence of the perturbation.

In mammalian cells, transcription factors within the same structural family commonly recognize similar DNA motifs but regulate diverse biological processes (Badis et al., 2009; Berger et al., 2008; Wei et al., 2010). It is unclear how faithful and distinct regulation is achieved with transcription factors of overlapping specificity. Our results suggest that chromatin structure influences the accessibility of potential binding sites and competition among transcription factors can facilitate discrimination of subtle differences in DNA binding specificity across the genome. Cooperation with conditionally, spatially or temporally expressed coactivators (Sharrocks, 2001) may provide additional selection for binding events to elicit specific transcriptional outcomes. Moreover, competition among transcription factors may have significant effects on transcriptional regulation, controlling the threshold for gene activation and limiting potential crosstalk between different signaling pathways. Overall, our findings suggest that the specificity of transcriptional regulation is a composite of the DNA binding specificity, nuclear abundance and functional interactions of each transcription factor in the context of eukaryotic genome architecture. Understanding the grammar of specific regulation of individual transcription factors will provide insights into decoding the complex regulatory network of the eukaryotic genomes.

EXPERIMENTAL PROCEDURES

Strains

Methods of strain construction are described in the Supplemental Experimental Procedures. All strains used in this study are listed in Table S1.

Media and Growth Conditions

Media and growth conditions are described in the Supplemental Experimental Procedures.

Defining the Consensus Pho4 Binding Motif

High-affinity Pho4 binding motifs were determined with the position specific scoring matrix (PSSM) described in Lam et al. (2008), which was derived from in vitro measurement of Pho4 DNA binding affinities (Maerkl and Quake, 2007). We selected the most stringent threshold (0.0075) to recapitulate in vivo validated high-affinity binding sites as the threshold of Pho4 high-affinity binding motifs (Lam et al., 2008). Since all determined high-affinity binding motifs at this threshold contain "CACGTG" as core sequence, we define these motifs as the consensus "CACGTG" binding motif for Pho4. All "NCACGTGN" motifs except "TCACGTGA" in *S. cerevisiae* genome meet the threshold.

Biotin-Tagging Immunoprecipitation with High-Throughput Sequencing

The procedure for biotin-tagging immunoprecipitation is described in detail in the Supplemental Experimental Procedures. In summary, we performed biotin-tagging immunoprecipitation with high-throughput sequencing (Bio-ChIP-seq) experiments on Pho2, Pho4, Cbf1, and mock samples in both high-and no-Pi conditions (60 min after Pi starvation), and on Pho4 in a *pho2Δ* strain grown in no-Pi conditions, and in *pho80Δ* and *cbf1Δ pho80Δ* strains grown in high-Pi conditions. Sequencing libraries were prepared for both ChIP DNA and input DNA (from the supernatant of total cell lysate) according to the Illumina protocol. Libraries with size between 200 and 300 bp were selected for PCR amplification and sequenced with an Illumina Genome analyzer II. Thirty-six base sequence tags were aligned to the *S. cerevisiae* genome with ELAND. On average, 2.9 and 12.5 million uniquely aligned reads were obtained for ChIP and input samples, respectively. Bio-ChIP-seq results of all Pho4 consensus binding sites are presented in Table S2.

In Vivo Nucleosome Mapping

Libraries from mononucleosomal DNA were sequenced from both ends with an Illumina genome analyzer II; details are described in the Supplemental Experimental Procedures. We assumed that the center of each sequenced DNA fragment is the nucleosome dyad and extended 73 bp on both sides to generate mononucleosome coverage (Figures S1C–S1E). We obtained nucleosome maps for the wild-type strain in high- and no-Pi conditions (40 min after Pi starvation) and for the *cbf1Δ* strain in high-Pi conditions. For each sample, 8–10 million uniquely aligned nucleosomal DNA sequencing reads were obtained. Nucleosome occupancy of all Pho4 consensus binding sites is presented in Table S2.

Microarray and Data Processing

Details of yeast cell collection, RNA isolation, complementary DNA (cDNA) synthesis and labeling, and microarray hybridization are described in the Supplemental Experimental Procedures. Whole-genome expression profiling was performed with competitive hybridization to Agilent 8 × 15K *S. cerevisiae* two-color expression microarrays (G2509F, AMADID #019838). Microarrays were scanned with an Axon 4000B scanner, and gene features were extracted with GenePix 5.1 software. Lowess and quantile normalization

were performed with the MATLAB bioinformatics toolbox before further analysis.

Wild-type (WT) no-Pi versus WT high-Pi, *cbf1Δ* high-Pi versus WT high-Pi, *cbf1Δ pho4Δ* high-Pi versus WT high-Pi, and *tye7Δ rtg3Δ* high-Pi versus WT high-Pi microarrays were performed with dye swaps to eliminate dye labeling bias and were analyzed in four biological replicates (Churchill, 2002; Yang and Speed, 2002). *pho80Δ* versus *pho80Δ pho4Δ* and *pho80Δ cbf1Δ* versus *pho80Δ cbf1Δ pho4Δ* were performed in high-Pi conditions and analyzed in two biological replicates. Mutant cycle analysis was constructed with a cyclic comparison so that the expression components could be directly inferred and dye labeling bias would be cancelled in the analysis (Churchill, 2002; Quackenbush, 2002; Yang and Speed, 2002) (Figure S4). The mutant cycle was repeated with three biological replicates.

Mutant Cycle Analysis, also known as Epistasis Expression Analysis

Details of the design and analysis of the mutant cycle (Figure S4) are described in the Supplemental Experimental Procedures. Eighty genes were selected as Pi starvation-responsive genes by direct comparison of the expression of the wild-type in no- and high-Pi conditions (activated more than 1.8-fold and significantly induced with a null hypothesis of < 1.5 fold, $p \leq 0.01$). For each of those genes, linear regression was performed and genes with significant Pho2, Pho4, or Co components (significantly induced with a null hypothesis of < 1.4 fold, $p \leq 0.05$) (Capaldi et al., 2008) are shown in Figure 4A. None of the genes induced after Pi starvation contains a significant Pho2 only component, indicating that Pho4 is the primary activator and Pho2 functions only as a cooperating factor (Komeili and O'Shea, 1999; Springer et al., 2003). Full mutant cycle analysis results are summarized in Table S3.

Determining Whether Binding Sites Are Accessible or Inaccessible

We calculate the average nucleosome occupancy of a 20 bp window centered on all consensus "CACGTG" sites to determine local nucleosome occupancy. Approximately 80% of the *S. cerevisiae* genome is estimated to be covered with nucleosomes (Lee et al., 2007); we thus say that a transcription factor binding site is in the accessible state if the average nucleosome occupancy on the site is in the lower quartile of the average genome nucleosome occupancy. Symmetrically, we say that the site is inaccessible if the average nucleosome occupancy is in the upper quartile of the average genome nucleosome occupancy. The average genome nucleosome occupancy distribution was determined by random sampling of the 20 bp window average for 10,000 genome locations and repetition of this ten times. The average of the upper and lower quartile cutoff for ten-time sampling of the average genome nucleosome occupancy was used to define the accessible and inaccessible states of binding sites. Varying the percentile threshold over a wide range (15%–30%) does not affect the outcome (Figure S2A).

Determining Binding at Consensus CACGTG Sites

Methods to determine the binding at Pho4 consensus binding sites are described in the Supplemental Experimental Procedures.

Pho2 and Pho4 Binding Cooperativity

The increase in Pho2 and Pho4 binding after Pi starvation (recruitment) is calculated with the following equation: taking Pho4 as an example, $\text{Enrichment}_{\text{No/High}} = (\text{Pho4}_{\text{ChIP No}} / \text{Pho4}_{\text{Input No}}) / (\text{Pho4}_{\text{ChIP High}} / \text{Pho4}_{\text{Input High}})$. To estimate a threshold to define transcription factor recruitment, we calculated the mean and standard deviation of Pho2 and Pho4 recruitment at all consensus "CACGTG" sites, excluding the sites at regulatory regions of Pho4 regulated genes (functional binding sites). We used mean + 2 standard deviations (SDs) as the threshold to identify the sites showing the most significant recruitment (Figure 4B). 23 of 28 sites showing recruitment of both Pho2 and Pho4 are regulated by both factors ($p = 8.1 \times 10^{-16}$, Fisher's exact test), and none of the Pho4-only recruited sites is regulated by Pho4 ($p = 0.003$, Fisher's exact test).

Pho4 Binding at Nonconsensus Binding Sites

Information about Pho4 binding at its nonconsensus binding sites is described in the Supplemental Experimental Procedures.

Fluorescence Microscopy

The procedures for fluorescence microscopy are described in the Supplemental Experimental Procedures.

Prediction of Pho4 Binding and Function

To predict Pho4 binding at its consensus binding motifs, we integrate information about Pho4 DNA binding preference, local nucleosome occupancy, and competition from Cbf1 into an equilibrium model (Granek and Clarke, 2005). At the equilibrium state, the probability of Pho4 not binding to an 8-mer sequence at position *i* can be expressed as

$$P_{unbound,\,i} = \frac{W_{Nuc,i} + W_{Cbf1,i} + 1}{W_{Nuc,i} + W_{Cbf1,i} + K_{a,\,i,\,adj} * [\text{Pho4}] + 1},$$

where $W_{Nuc,\,i}$, $W_{Cbf1,\,i}$, $K_{a,\,i,\,adj}$ and [Pho4] represent occlusion from nucleosomes, competition from Cbf1, equilibrium association constant of Pho4 with the 8-mer sequence, and the nuclear Pho4 concentration.

If there are *n* potential Pho4 binding sites near position *x* on the genome, the binding probability at *x* is finally calculated as

$$P_{bound,\,x} = 1 - \Pi\, P_{unbound,\,i} (i = 1, 2, 3 \ldots n).$$

Derivation of equations and details about the calculation are described in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

Agilent custom 8 × 15K S. cerevisiae two-color whole genome expression microarray data are deposited in the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE23580. All sequencing data are deposited in the GEO database under accession number GSE29506.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and three tables and can be found with this article online at doi:10.1016/j.molcel.2011.05.025.

ACKNOWLEDGMENTS

We thank C. Daly and J. Zhang for help with Illumina sequencing; X. Zhang for assistance with microscopy; former and present O'Shea lab members for discussion and commentary; and V. Denic, N. Friedman, A. Regev, and B. Stern for critical reading of the manuscript. This work was supported by National Institutes of Health grant R01 GM051377 and the Howard Hughes Medical Institute.

Received: October 7, 2010
Revised: January 8, 2011
Accepted: May 18, 2011
Published: June 23, 2011

REFERENCES

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.

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., et al. (2009). Diversity and complexity in DNA recognition by transcription factors. Science *324*, 1720–1723.

Barbarić, S., Münsterkötter, M., Svaren, J., and Hörz, W. (1996). The homeodomain protein Pho2 and the basic-helix-loop-helix protein Pho4 bind DNA cooperatively at the yeast PHO5 promoter. Nucleic Acids Res. *24*, 4479–4486.

Barbaric, S., Münsterkötter, M., Goding, C., and Hörz, W. (1998). Cooperative Pho2-Pho4 interactions at the PHO5 promoter are critical for binding of Pho4 to UASp1 and for efficient transactivation by Pho4 at UASp2. Mol. Cell. Biol. *18*, 2629–2639.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Peña-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., et al. (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell *133*, 1266–1276.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., et al; ENCODE Project Consortium; NISC Comparative Sequencing Program; Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center; Broad Institute; Children's Hospital Oakland Research Institute. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799–816.

Capaldi, A.P., Kaplan, T., Liu, Y., Habib, N., Regev, A., Friedman, N., and O'Shea, E.K. (2008). Structure and function of a transcriptional network activated by the MAPK Hog1. Nat. Genet. *40*, 1300–1306.

Churchill, G.A. (2002). Fundamentals of experimental design for cDNA microarrays. Nat. Genet. Suppl. *32*, 490–495.

Farnham, P.J. (2009). Insights from genomic profiling of transcription factors. Nat. Rev. Genet. *10*, 605–616.

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.

Granek, J.A., and Clarke, N.D. (2005). Explicit equilibrium modeling of transcription-factor binding and gene regulation. Genome Biol. *6*, R87.

Graumann, J., Dunipace, L.A., Seol, J.H., McDonald, W.H., Yates, J.R., 3rd, Wold, B.J., and Deshaies, R.J. (2004). Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast. Mol. Cell. Proteomics *3*, 226–237.

Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.B., Reynolds, D.B., Yoo, J., et al. (2004). Transcriptional regulatory code of a eukaryotic genome. Nature *431*, 99–104.

Hochheimer, A., and Tjian, R. (2003). Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. Genes Dev. *17*, 1309–1320.

Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686–691.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. *3*, 318–356.

Jimenez-Sanchez, G., Childs, B., and Valle, D. (2001). Human disease genes. Nature *409*, 853–855.

Kasowski, M., Grubert, F., Heffelfinger, C., Hariharan, M., Asabere, A., Waszak, S.M., Habegger, L., Rozowsky, J., Shi, M., Urban, A.E., et al. (2010). Variation in transcription factor binding among humans. Science *328*, 232–235.

Kent, N.A., Tsang, J.S., Crowther, D.J., and Mellor, J. (1994). Chromatin structure modulation in Saccharomyces cerevisiae by centromere and promoter factor 1. Mol. Cell. Biol. *14*, 5229–5241.

Kent, N.A., Eibert, S.M., and Mellor, J. (2004). Cbf1p is required for chromatin remodeling at promoter-proximal CACGTG motifs in yeast. J. Biol. Chem. *279*, 27116–27123.

Khorasanizadeh, S. (2004). The nucleosome: from genomic organization to genomic regulation. Cell *116*, 259–272.

Kolodziej, K.E., Pourfarzad, F., de Boer, E., Krpic, S., Grosveld, F., and Strouboulis, J. (2009). Optimal use of tandem biotin and V5 tags in ChIP assays. BMC Mol. Biol. *10*, 6.

Komeili, A., and O'Shea, E.K. (1999). Roles of phosphorylation sites in regulating activity of the transcription factor Pho4. Science *284*, 977–980.

Kornberg, R.D., and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell *98*, 285–294.

Lam, F.H., Steger, D.J., and O'Shea, E.K. (2008). Chromatin decouples promoter threshold from dynamic range. Nature *453*, 246–250.

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.

Lee, T.A., Jorgensen, P., Bognar, A.L., Peyraud, C., Thomas, D., and Tyers, M. (2010). Dissection of combinatorial control by the Met4 transcriptional complex. Mol. Biol. Cell *21*, 456–469.

Liu, X., Lee, C.K., Granek, J.A., Clarke, N.D., and Lieb, J.D. (2006). Whole-genome comparison of Leu3 binding in vitro and in vivo reveals the importance of nucleosome occupancy in target site selection. Genome Res. *16*, 1517–1528.

MacIsaac, K.D., Wang, T., Gordon, D.B., Gifford, D.K., Stormo, G.D., and Fraenkel, E. (2006). An improved map of conserved regulatory sites for Saccharomyces cerevisiae. BMC Bioinformatics *7*, 113.

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

Narlikar, G.J., Fan, H.Y., and Kingston, R.E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell *108*, 475–487.

O'Neill, E.M., Kaffman, A., Jolly, E.R., and O'Shea, E.K. (1996). Regulation of PHO4 nuclear localization by the PHO80-PHO85 cyclin-CDK complex. Science *271*, 209–212.

Ogawa, N., DeRisi, J., and Brown, P.O. (2000). New components of a system for phosphate accumulation and polyphosphate metabolism in Saccharomyces cerevisiae revealed by genomic expression analysis. Mol. Biol. Cell *11*, 4309–4321.

Pan, Y., Tsai, C.J., Ma, B., and Nussinov, R. (2010). Mechanisms of transcription factor selectivity. Trends Genet. *26*, 75–83.

Pierce, M., Benjamin, K.R., Montano, S.P., Georgiadis, M.M., Winter, E., and Vershon, A.K. (2003). Sum1 and Ndt80 proteins compete for binding to middle sporulation element sequences that control meiotic gene expression. Mol. Cell. Biol. *23*, 4814–4825.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569–577.

Quackenbush, J. (2002). Microarray data normalization and transformation. Nat. Genet. Suppl. *32*, 496–501.

Robinson, K.A., and Lopes, J.M. (2000). SURVEY AND SUMMARY: Saccharomyces cerevisiae basic helix-loop-helix proteins regulate diverse biological processes. Nucleic Acids Res. *28*, 1499–1505.

Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S., et al. (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science *328*, 1036–1040.

Schneider, K.R., Smith, R.L., and O'Shea, E.K. (1994). Phosphate-regulated inactivation of the kinase PHO80-PHO85 by the CDK inhibitor PHO81. Science *266*, 122–126.

Sharrocks, A.D. (2001). The ETS-domain transcription factor family. Nat. Rev. Mol. Cell Biol. *2*, 827–837.

Springer, M., Wykoff, D.D., Miller, N., and O'Shea, E.K. (2003). Partially phosphorylated Pho4 activates transcription of a subset of phosphate-responsive genes. PLoS Biol. *1*, E28.

van Werven, F.J., and Timmers, H.T. (2006). The use of biotin tagging in Saccharomyces cerevisiae improves the sensitivity of chromatin immunoprecipitation. Nucleic Acids Res. *34*, e33.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. *10*, 252–263.

Vogel, K., Hörz, W., and Hinnen, A. (1989). The two positively acting regulatory proteins PHO2 and PHO4 physically interact with PHO5 upstream activation regions. Mol. Cell. Biol. *9*, 2050–2057.

Wasson, T., and Hartemink, A.J. (2009). An ensemble model of competitive multi-factor binding of the genome. Genome Res. *19*, 2101–2112.

Wei, G.H., Badis, G., Berger, M.F., Kivioja, T., Palin, K., Enge, M., Bonke, M., Jolma, A., Varjosalo, M., Gehrke, A.R., et al. (2010). Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. EMBO J. *29*, 2147–2160.

Yang, Y.H., and Speed, T. (2002). Design issues for cDNA microarray experiments. Nat. Rev. Genet. *3*, 579–588.

Zheng, W., Zhao, H., Mancera, E., Steinmetz, L.M., and Snyder, M. (2010). Genetic analysis of variation in transcription factor binding in yeast. Nature *464*, 1187–1191.

Zhu, C., Byers, K.J., McCord, R.P., Shi, Z., Berger, M.F., Newburger, D.E., Saulrieta, K., Smith, Z., Shah, M.V., Radhakrishnan, M., et al. (2009). High-resolution DNA-binding specificity analysis of yeast transcription factors. Genome Res. *19*, 556–566.