Modeling the regulatory network of histone acetylation in *Saccharomyces cerevisiae*

Hung Pham[1], Roberto Ferrari[2], Shawn J Cokus[1], Siavash K Kurdistani[2] and Matteo Pellegrini[1,*]

[1] Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA, USA and [2] Department of Biological Chemistry, University of California, Los Angeles, CA, USA
* Corresponding author. Department of Molecular, Cell and Developmental Biology, 621 Charles Young Drive South, University of California, Los Angeles, CA 90095, USA.
Tel.: +1 310 825 0012; Fax: +1 310 206 3987; E-mail: matteop@mcdb.ucla.edu

Acetylation of histones plays an important role in regulating transcription. Histone acetylation is mediated partly by the recruitment of specific histone acetyltransferases (HATs) and deacetylases (HDACs) to genomic loci by transcription factors, resulting in modulation of gene expression. Although several specific interactions between transcription factors and HATs and HDACs have been elaborated in *Saccharomyces cerevisiae,* the full regulatory network remains uncharacterized. We have utilized a linear regression of optimized sigmoidal functions to correlate transcription factor binding patterns to the acetylation profiles of 11 lysines in the four core histones measured at all *S. cerevisiae* promoters. The resulting associations are combined with large-scale protein–protein interaction data sets to generate a comprehensive model that relates recruitment of specific HDACs and HATs to transcription factors and their target genes and the resulting effects on individual lysines. This model provides a broad and detailed view of the regulatory network, describing which transcription factors are most significant in regulating acetylation of specific lysines at defined promoters. We validate the model, both computationally and experimentally, to demonstrate that it yields accurate predictions of these regulatory mechanisms.
Molecular Systems Biology 18 December 2007; doi:10.1038/msb4100194
Subject Categories: chromatin & transcription
Keywords: histone acetylation; linear regression; networks; regulatory network

Introduction

Transcriptional regulation is a highly complex process that involves not only the recruitment of polymerases to active promoters but also the remodeling of chromatin. Within silent chromatin, DNA is tightly wrapped around histone proteins (Richmond and Davey, 2003) leading to transcriptional suppression. In contrast, in active promoters, the chromatin state is more open, enabling access to the transcription machinery. One of the mechanisms used to control the state of chromatin structure is the modification of histones (Kurdistani and Grunstein, 2003). Histone acetyltransferases (HATs) and histone deacetylases (HDACs) alter the charge state of lysines in histone tails by adding and removing acetyl groups. The pattern of these modifications, along with methylation, phosphorylation and ubiquitination, contributes to establish a proposed histone code that regulates transcription through alterations of chromatin structure and generation of binding sites for chromatin-interacting factors (Strahl and Allis, 2000; Jenuwein and Allis, 2001).

Histone acetylation patterns at individual genomic loci are in large part determined by the recruitment of specific HATs and HDACs by transcription factors (TFs) (Kurdistani and Grunstein, 2003). For instance, the Ume6 TF recruits the Rpd3 HDAC to its target genes, leading to repression of the downstream genes (Rundlett *et al*, 1998). Although a few interactions between HATs and HDACs and specific TFs have been established, we do not yet know the full network of interactions between these and the effect that such interactions have on establishing acetylation patterns on histone tails. While TF binding has been correlated with histone acetylation patterns, a regulatory network that controls histone acetylation levels has not been determined (Guo *et al*, 2006). Therefore, we set out to utilize genome-wide TF-binding and acetylation data, along with protein–protein interaction data, to determine this network on a genome-wide scale.

To construct a comprehensive network, we utilized two data sets that provide a systematic measurement of the acetylation levels of 11 lysines (Kurdistani *et al*, 2004) and the binding

levels of 204 TFs in promoter regions (Harbison *et al*, 2004) on a genome-wide basis in *Saccharomyces cerevisiae*. In addition, protein–protein interaction data from all yeast complexes have been recently determined (Gavin *et al*, 2006), and have been added to the pre-existing set of interactions contained in various databases such as DIP (Salwinski *et al*, 2004) and SGD. Using multiple linear regression of optimized sigmoidal functions, we construct a network that links TFs to individual lysines in histones. When known, we also introduce in our model interactions between TFs and specific histone-modifying enzymes. We provide both computational and experimental validation of the network, indicating that it accurately captures the regulation of acetylation levels of histones promoters in *S. cerevisiae* on a genome-wide scale.

We find that there is selectivity of histone acetylation site usage by various TFs, with each TF associating with a specific subset of acetylation sites. We also find that certain acetylation sites are linked to a greater number of TFs than others, indicating that some acetylation sites may have widespread roles in transcription. We show that the network supports the view that the deacetylation of promoters tends to repress transcription, and acetylation to activate it, although the regulation of histone genes represents an interesting exception to this rule. Finally, we show that in most cases the TFs cause the changes in acetylation through the recruitment of HATs and HDACs, rather than being recruited by the histone modifications.

Results

Construction of a global histone acetylation network

We have constructed a model of the regulatory network of histone acetylation in *S. cerevisiae*. This model describes the interactions between HAT and HDAC complexes, the TFs that recruit them and the lysines that are modified. To construct this model, we analyzed two data sets that measure the binding of TFs and the acetylation state of histones at all yeast promoters. Both data sets were colleted in rich media; therefore our model represents the regulatory network in this condition and may potentially be altered in other conditions that subject the cell to stresses. Since our hypothesis is that the particular acetylation profile of a promoter is in large part determined through the recruitment of HATs and HDACs by TFs that bind the promoter, we constructed a multivariate linear regression framework of optimized sigmoidal functions for modeling acetylation in terms of TF-binding data.

Multivariate regression allows us to model combinatorial regulation by multiple TFs. That is, the acetylation level of a gene may be determined by the binding of multiple factors to its promoter. In fact, the majority of the promoters in our model are regulated by more than one TF, as shown in Supplementary Figure 1. Nonlinear sigmoidal functions allow us to model the expected sigmoidal relationship between TF-binding and acetylation levels (see Supplementary Figure 2). If the binding of a TF to a promoter, as measured by the logarithm of the ratios of immunoprecipitated DNA to input, is

negative (implying little or no binding), the effect on acetylation is close to zero, whereas if the binding is positive and large, the effect saturates at some maximal level of acetylation. We use the midpoint of the sigmoidal function as the threshold for defining targets of a factor: promoters that are bound by levels above the midpoint are considered bona fide targets, whereas those with lower binding are not. The distribution of the number of targets per factor is shown in Supplementary Figure 3 and shows that the factors may generally be classified as global regulators that affect many genes, or local ones that affect a small number (less than 200).

Using this approach, we describe the acetylation of each lysine as a linear superposition of the binding profiles of all TFs transformed by each function. For example, we model the acetylation level of H3K18ac at each promoter i as follows:

$$H3K18ac_i = \sum_{j=1}^{N} F_j(B_{ij})$$

where F_j is the sigmoidal function that describes the coupling between TF j and H3K18ac, and B_{ij} is the binding of TF j to promoter i. The functional form of F is fit for each TF, but is kept constant for a specific factor across all promoters. Thus we need to fit three parameters per TF in our model: the first parameter describes the midpoint of the sigmoidal function, the second its slope and the third its maximal value. If the maximal value of the function is positive, then we hypothesize that the binding of the factor leads to the acetylation of the promoter. In contrast, if the maximal value is negative, then the binding leads to deacetylation. Multivariate regression identifies the most significant TFs in the model and the functional form F between each TF and histone lysine. These relationships may then be represented as a network that contains the acetylated lysines and the significant factors that modify them. The details of this approach are described in Materials and methods below.

The full network of the relationships between TFs and histone lysines is shown in Figures 1 and 2. For clarity, we have separated the full network into two subnetworks that naturally emerge from this analysis: one network describes the regulation of hyperacetylation while the other hypoacetylation. Both networks depict the 11 histone lysines whose acetylation levels are measured at all promoters (Kurdistani *et al*, 2004) as yellow ovals. The TFs that emerge in our model as significant regulators of the acetylation or deacetylation of specific lysines are depicted as rounded rectangles in shades of blue or red, indicating the average magnitude of the maximal value of the associated sigmoidal function; dark blue indicates a large positive maximal value while dark red a large negative one. The shade of the link indicates the maximal value of the sigmoidal function that couples a specific TF and lysine—blue for positive values and red for negative ones.

We hypothesize that these TFs recruit proteins that are responsible for the acetylation and deacetylation of histone lysines. As a result, we have extended the network by including interactions between HATs and HDACs and TFs from protein–protein interactions databases. These are shown in the second panel on the right side of the networks. Finally, to complete the network, we show the known HAT and HDAC complexes in the rightmost panel.
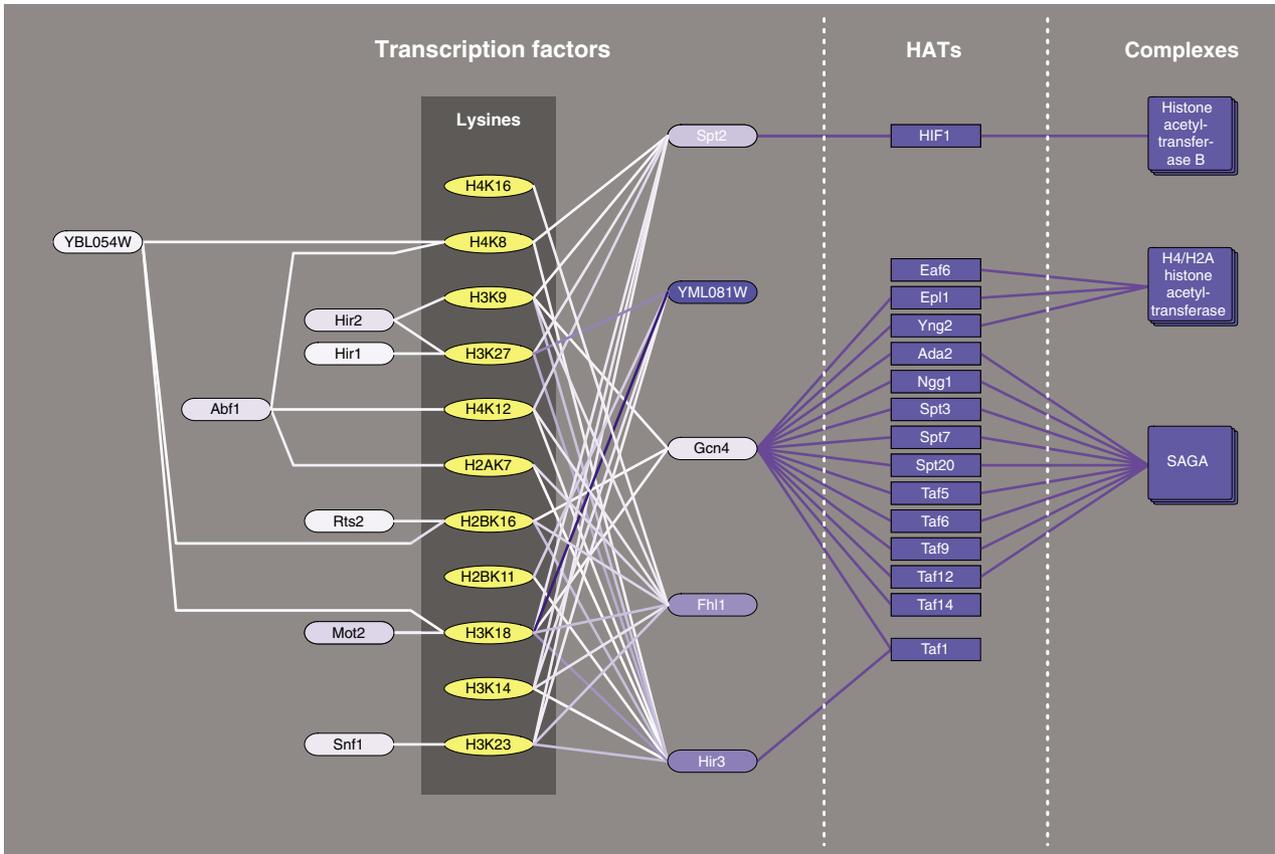
Figure 1 The regulatory network of histone acetylation in *Saccharomyces cerevisiae*. The network has been divided into two subnetworks: the one shown here describes the regulation of histone hyperacetylation. The yellow ovals represent the 11 lysines whose acetylation at each promoter has been measured. The TFs that we predict to regulate the acetylation levels of these lysines are shown linked to the lysines. The color of the TFs and the links between TFs and lysines are determined by the linear regression coefficients. Blue indicates correlation with hyperacetylation. White indicates a weak association. The known HAT and HDAC complexes, along with their subunits, are also included in the network. The interactions between HATs, HDACs and TFs are determined from protein–protein interaction data.

Combinatorial interaction of TFs and histone acetylation

Considering that there are more TFs than histone modifiers or histone acetylation sites, there should be overlapping but distinct patterns of acetylation site usage by TFs. The network shows this to be indeed the case. Figure 3A, derived from the combination of the two hyper- and hypoacetylation networks, shows the number of TFs that are associated with each lysine in our network. From the number of interactions, we identify H3K18 as the most widely regulated acetylation site with links to 15 TFs. At the opposite end of the spectrum, H4K16 is regulated by only two TFs.

It is interesting to note that H3K18 and H4K16 show the least correlation between any pair of acetylated lysines, which are otherwise often highly correlated. H3K18 acetylation appears to be a general mark of active transcription (Kurdistani *et al*, 2004). In contrast, H3K16 could be a more specific mark, possibly associated with the regulation of ribosomal genes as suggested by the fact that FHL1, a known regulator of rRNA processing, is one of the few factors that appears to regulate it. These findings suggest that certain acetylation sites, by virtue of being regulated by multiple TFs, may have broader and

more general roles in transcriptional regulation while others are likely to have more specific ones.

Figure 3B describes the converse relationship: the number of lysines to which each TF is linked. Here, we find a more dramatic distribution, with five TFs interacting with nine or more lysines in class 1, while nine TFs interact with only one lysine in class 2. We investigated whether there were general trends that separated the two classes of TFs. We found that in both hyper- and hypoacetylation networks, TFs that interact with many lysines are more likely to have direct interactions with HATs and HDACs (total eight) than those that interact with single lysines (total zero), possibly indicating that class 1 TFs have stronger interactions with the HAT or HDAC complexes, leading to the modification of more lysines.

Validation of the network

The network demonstrates the specificity of regulation of histone acetylation within cells. Of the 204 TFs analyzed, we are able to identify 26 TFs that are likely to regulate histone acetylation at their target genes. These include factors that have previously been associated with this role as well as new factors that were not known to function in regulating
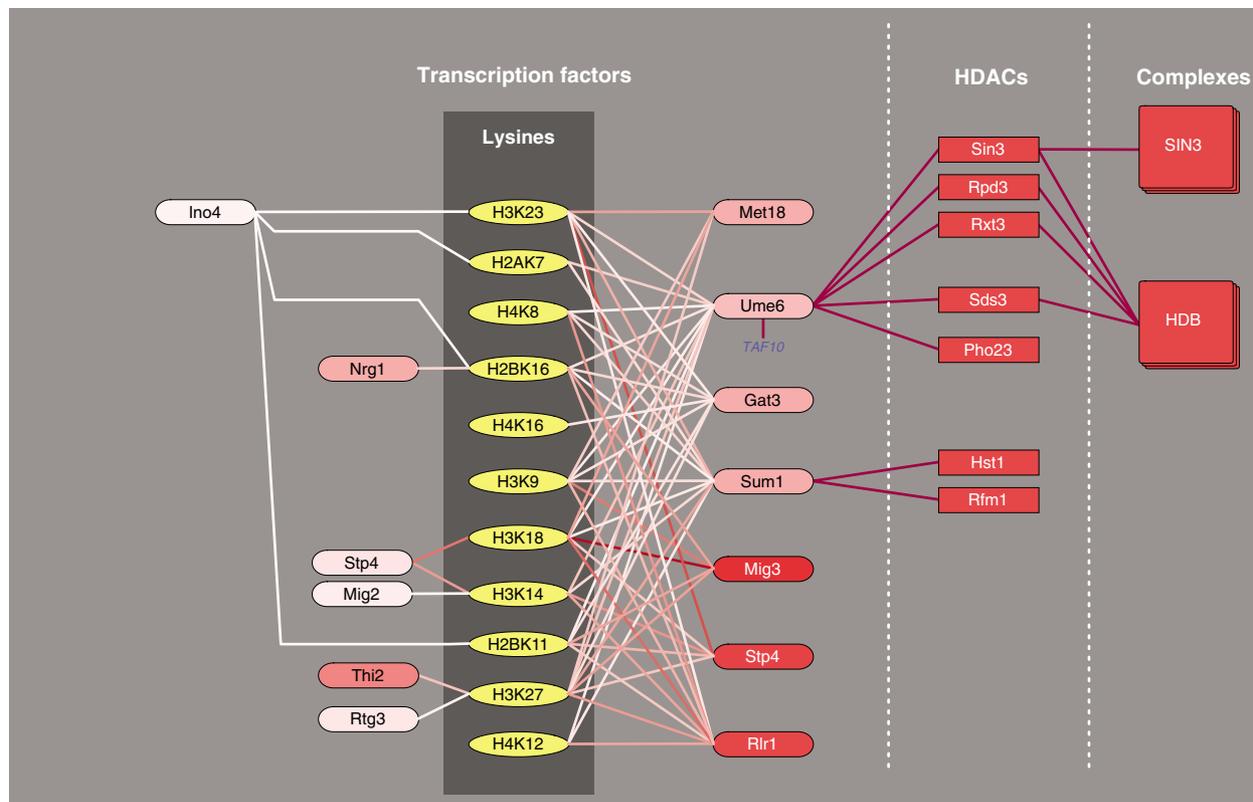
Figure 2 The regulatory network of histone acetylation in *Saccharomyces cerevisiae*. The network has been divided into two subnetworks: the one shown here describes the regulation of histone hypoacetylation. The yellow ovals represent the 11 lysines whose acetylation at each promoter has been measured. The TFs that we predict to regulate the acetylation levels of these lysines are shown linked to the lysines. The color of the TFs and the links between TFs and lysines are determined by the linear regression coefficients. Red nodes and links indicate that the TF is correlated with hypoacetylation. White indicates a weak association. The known HAT and HDAC complexes, along with their subunits, are also included in the network. The interactions between HATs, HDACs and TFs are determined from protein–protein interaction data.

acetylation levels. In both cases, we have included in the network the known interactions between TFs and subunits of the HAT and HDAC complexes. As a result, the model generates specific axes of interaction between histone modifiers, TFs at their target genes and specific lysine residues.

We validate the couplings between TFs and lysines in our network using two approaches: through a statistical test and by verifying individual relationships experimentally. The statistical validation is based on the assumption that TFs that are correlated with low acetylation levels (red) should interact with HDACs. Conversely, we expect that TFs that are correlated with high acetylation levels should interact with HATs. We find that our network contains 22 correct associations and only one incorrect one. This is striking since our TF–lysine interactions are not trained on protein–protein interactions at all, yet we are able to correctly identify the type of nearly all the TF–HAT or HDAC interactions. We performed a statistical test to measure how often we obtain these correct associations versus incorrect ones in the real versus random networks and found that the number of correct associations in the real network is significant below the 1e-5 level.

We have also conducted experimental validation of the network. We tested the regulatory interactions described in our network by measuring changes in acetylation in mutant strains of *S. cerevisiae* that lack specific TFs. In the TF mutant

strains, the recruitment of the HATs or HDACs that are responsible for the observed acetylation should be eliminated, yielding a corresponding change in acetylation, which we measured. We tested six TFs, three that were significant in our model, Hir3, Sum1 and YML081W, and three negative controls. Two of the three significant factors were selected because one interacted with a HAT and one with an HDAC; thus, we could test whether these had opposite effects on acetylation after deletion. The third had no interaction with HAT or HDAC enzymes and is virtually uncharacterized in the literature and therefore represents a potentially novel finding of our model. The three negative controls were factors that did not appear in any of our models, and we would therefore predict that their deletion has little or no effect on acetylation. This is in fact by and large what we see in Figure 4F and G, where the changes induced by the deletion of these control factors are generally small.

The acetylation network reveals specific interactions

The first experiment we conducted involved deleting Sum1, a transcriptional repressor required for mitotic repression of middle sporulation-specific genes (Xie *et al*, 1999). This factor
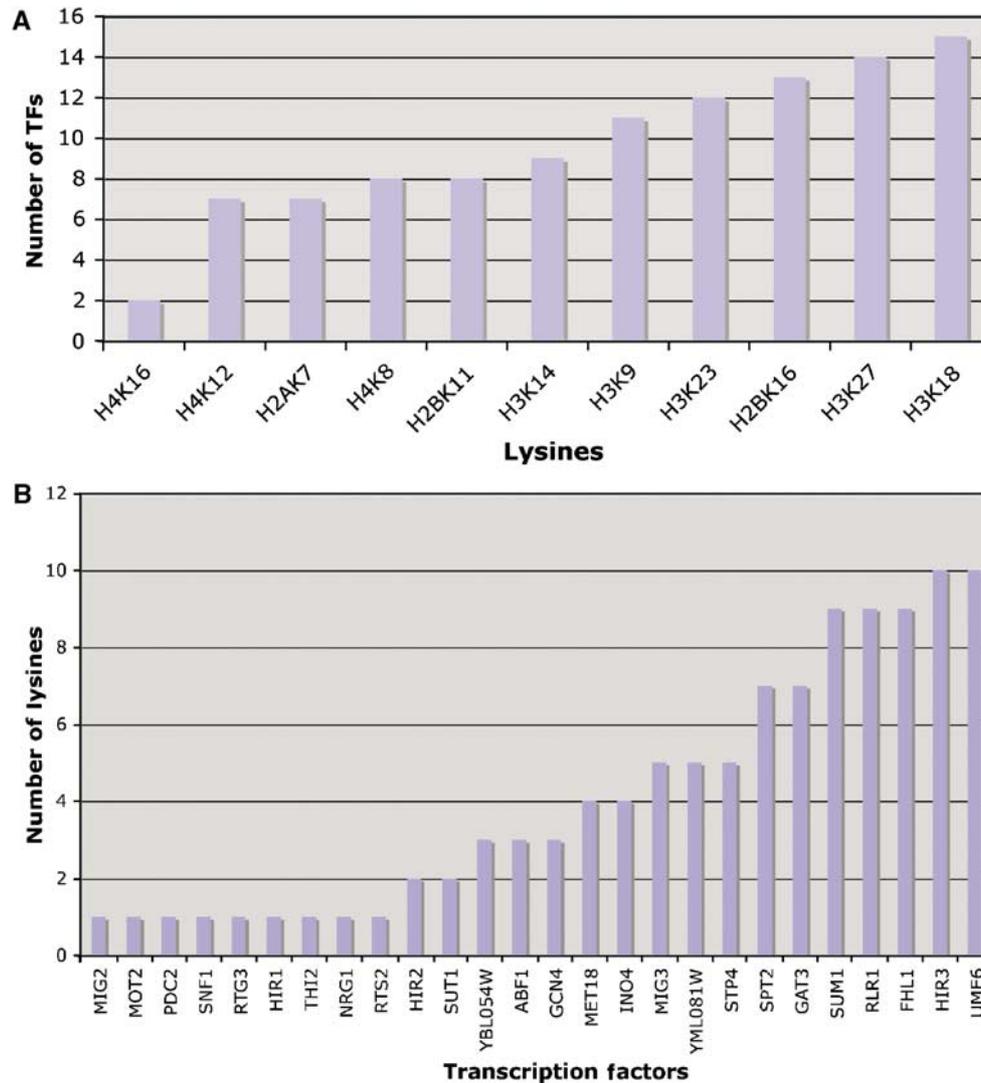
Figure 3 In the top panel we show that among all 11 acetylation sites, lysine H3K18 is linked to the largest number of TFs. This suggests that this lysine may play a more general role in regulating transcription than other lysines that are likely to regulate specific processes, such as H4K16. In the bottom panel we show the number of lysines that are linked to each TF. Ume6 and Hir3 are linked to all 10 lysines, while a large number of factors are linked to only one lysine.

is known to interact with Hst1, an NAD + -dependent histone deacetylase that is a subunit of the Sum1p/Rfm1p/Hst1p complex, and we would therefore expect it to regulate histone deacetylation. However, the specific lysines that it regulates and the targeted promoters were previously not known on a genomic scale. Our model allows us to define both the regulated lysines and the promoters that are regulated by Sum1 (based on the midpoint of its sigmoidal function). Accordingly, we predict that Sum1 regulates all lysines other than H4K16 and H2BK11 and it does so at ∼40 promoters, where the binding ratio is greater than two.

We tested these predictions by comparing acetylation levels of three lysines at three promoters between wild-type and sum1-deletion strains. At the first two promoters (iYFL011W and iYIR027C), the binding of Sum1 is above the threshold set by the sigmoidal function for all three lysines (H3K14, H4K16 and H3K18). The model predicts that both H3K14 and H3K18 levels are regulated by Sum1, whereas H4K16 levels are not. What we find at these two promoters is that the first two

lysines are in fact significantly hyperacetylated in the mutant with respect to the wild type, especially in the iYIR027C promoter, while K4K16 levels are not (see Figure 4A and B). This result supports our previous finding that the regulation of H4K16 is largely independent of H3K18. The third promoter we selected, iYDR224C, is one where Sum1 is bound below threshold, and we would therefore predict that the deletion of Sum1 should have little or no effect on acetylation. Again, the experiments largely support this prediction, since we see little or negative change in the acetylation levels of all three lysines at this promoter (Figure 4C).

Regulation of transcription by histone acetylation

It is generally assumed that the deacetylation of promoters leads to transcriptional repression, while the acetylation to activation. This may be in part due to the tightening and loosening of histone DNA interactions, as well as due to the recruitment of chromatin-modifying proteins that recognize
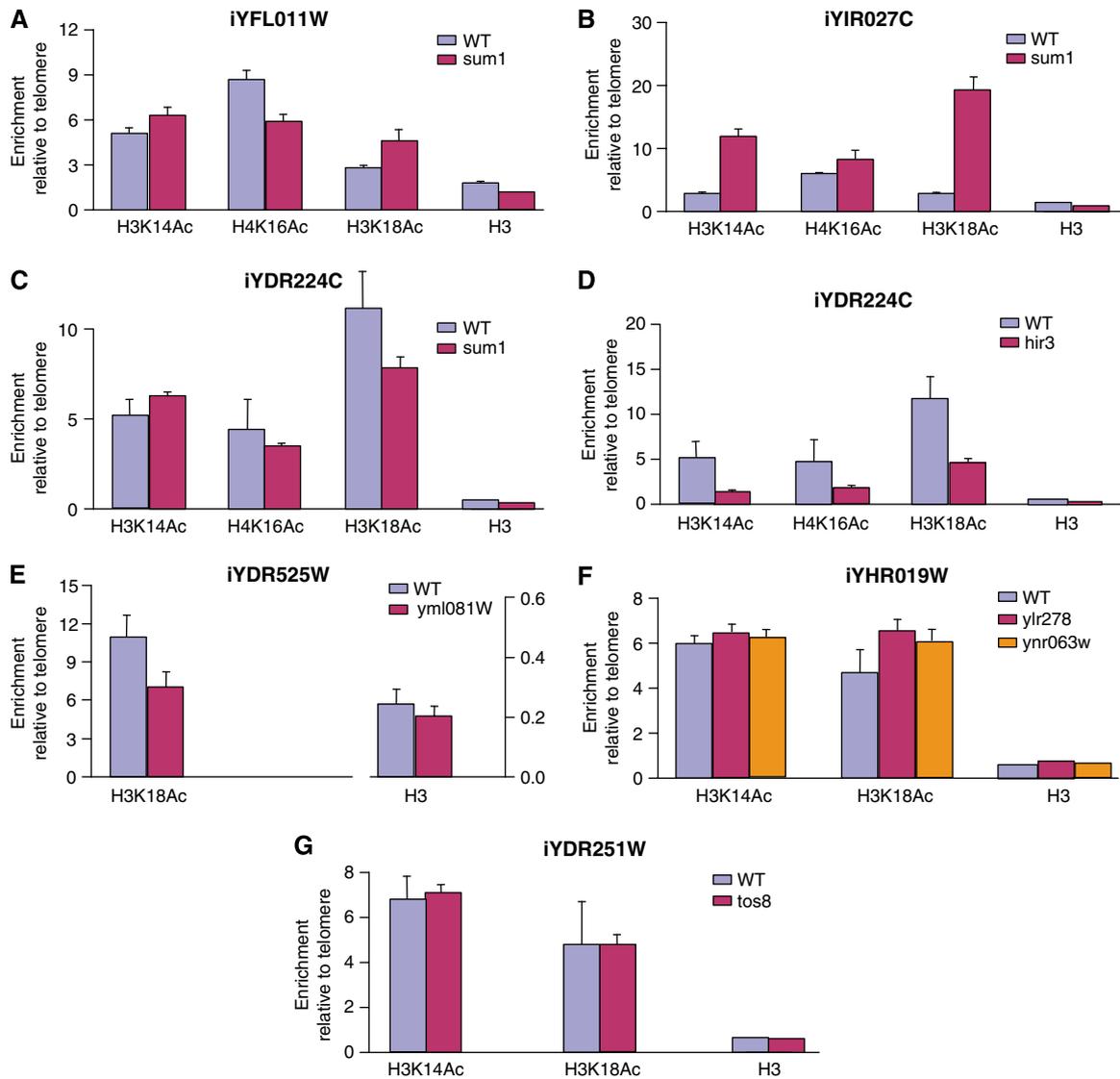
Figure 4 Experimental validation of the regulatory network. We examine the change in acetylation in six mutants that lack a TF. For each mutant, we measure the change in acetylation in one or more promoters where the factor is known to bind. In (**A**) and (**B**) we show that H3K14 and H3K18 acetylation levels increase in sum1 mutants at the YFL011W and YIR027C promoters where Sum1 binds strongly, but as shown in (**C**) are not significantly changed at the YDR244C promoter where Sum1 binds weakly. As predicted by our model H4K16 levels are not significantly affected by Sum1. In (**D**), we see that the hir3 mutant has significantly less acetylation at the YDR224C promoter that regulates the H2B gene. In (**E**), we see that Yml081W affects the acetylation of H3K18 at a promoter where it is bound. Finally, in (**F**) and (**G**), we see our control experiments, which show only small changes in promoters where factors that are not in our model are strongly bound.

specific acetylation marks. We tested whether our network supports this view of transcriptional regulation by annotating factors as transcriptional activators and repressors based on an analysis of the transcriptional profiles of knockout strains measured in rich media (Hu *et al*, 2007). As discussed in the Materials and methods section, we used the same linear regression of optimized sigmoidal function strategy to model expression profiles in terms of TF-binding profiles as we did for acetylation profiles. If the model of a specific profile contained the TF that was deleted in that mutant, then we were able to infer whether the factor was an activator or repressor, based on the sign of the maximal value of its sigmoidal function. If the sign was positive, and the genes bound by the deleted factor were overexpressed, then the factor was considered a

repressor. In contrast, if the sign was negative, then it was considered an activator because the genes it bound to were repressed in the mutant strain.

We computed models of the expression profiles for the 26 TFs found in our acetylation network and found five cases where the models contained the deleted factor (Table I). For many of the other cases, the mutant strain had a very similar expression profile to the wild type, or the perturbation was explained by the activation of other factors. From this analysis, we were able to infer that Gcn4 and Mot2 are activators, while Hir3, Sum1 and Ume6 are repressors, as one might expect from previous characterizations of these factors. Both the activators are associated with hyperacetylation in our network, while the repressors, with the exception of Hir2, are associated with

Table I Association of acetylation with transcriptional regulation

TF	Coef	Type	Recruits
GCN4	−2.1	Activator	HAT
MOT2	−0.2	Activator	HAT
UME6	1.4	Repressor	HDAC
HIR2	3.0	Repressor	HAT
SUM1	4.6	Repressor	HDAC

hypoacetylation. Thus in four of the five cases, our analysis supports the conventional view that transcriptional deacetylation of promoters leads to transcriptional repression, while acetylation leads to activation.

An intriguing exception to this rule is the regulation of histone genes. The Hir1, Hir2 and Hir3 TFs, which form the Hir complex, are believed to be primarily responsible for repressing the expression of these genes (Prochasson *et al*, 2005). This repression is usually removed in the G1/S phase of the cell cycle, when histone genes are transcribed at high levels. However, it appears that the Hir complex remains bound to the histone genes throughout the cell cycle, and therefore the exact mechanism of the switch that leads from repression to activation of these genes is not known. One possibility is that histone acetylation plays an important role in this transition, as the acetylation of certain lysines has been shown to vary in a cell cycle-dependent manner (Xu *et al*, 2005).

We tested the effect of Hir3 deletion on the acetylation levels of a histone H2B gene, YDR224C. Protein–protein interaction data indicate that Hir3 interacts with Taf1, a putative histone acetyltransferase. At the iYDR224C promoter region, deletion of hir3 leads to a marked decrease of acetylation at all three lysines that we examined, as predicted by our model (Figure 4D). This result then raises the possibility that the repressive function of the Hir complex is mediated by the recruitment of a histone acetyltrasferase. In the mutant strain, where the recruitment is inhibited, the promoters are no longer acetylated, leading to increased transcription of histone genes. Therefore, in contrast to the standard view, in this case acetylation of promoters appears to be associated with the repression of transcription and not the activation. Nonetheless, it does not appear that the acetylation levels of H3K18, one of the lysines we examined, vary periodically during the cell cycle (Xu *et al*, 2005). Therefore, although our data indicate that Hir3 regulates the acetylation of various lysines and is associated with the repression of histone genes, the mechanism by which this repression is relieved during the S phase remains to be determined.

The network predicts novel interactions for uncharacterized TFs

The network enables us to infer partial functions for previously uncharacterized TFs. For example, TF Yml081w is very poorly characterized in the literature. However, we find that it is significantly linked to several lysines in the hyperacetylation network and would therefore conclude that it likely recruits a HAT. To test this hypothesis, we constructed

strains lacking Yml081w and measured the change in acetylation at a promoter where the factor binds strongly. The result, shown in Figure 4E, indicates that there is a small but reproducible decrease in acetylation in the mutant strain, supporting the notion that this factor does in fact recruit a HAT.

This result leads us to postulate that Yml081w is likely an activator of transcription. To test this hypothesis, we looked at the expression of genes during the yeast cell cycle (Spellman *et al*, 1998). We compared the average expression of 153 genes strongly bound by this TF (log ratio greater than one) to the expression of the transcript of Yml081w. We find that the two are related by a correlation coefficient of 0.7, indicating that when the transcripts of Yml081w are high, the expression of its target genes is also high and vice versa. This finding therefore supports the hypothesis that Yml081w is likely an activator. Therefore, although our regulatory network cannot determine the specific function of each individual TF, as this example shows, in some cases it may allow us to predict whether a factor is an activator or repressor.

TFs recruit HATs and HDACs

An important question that emerges from our network is whether there are cases in which the histone modification recruits the TFs, as opposed to the factors recruiting HATs and HDACs. The analysis so far identifies only correlations between TFs and acetylation profiles, but is not able to establish a causal connection. To answer this question, we examined a data set in which four lysines were mutated to arginine to inhibit their acetylation (Dion *et al*, 2005). Each of the 15 possible combinatorial combinations of mutants was constructed and the resulting expression was measured. We again built models for the expression profiles as above and tested whether the presence of TFs in our model correlated with the mutant state of any of the four lysines.

We found that out of all possible combinations of TFs and lysines, only one was statistically significant: the factor Ste12 was present with a negative coefficient only in models in which H4K16 was mutated ($P < 0.001$). This could imply that Ste12 binding recognizes the acetylation of H4K16. However, Dion *et al* speculate that an alternate explanation is that the H4K16 mutant derepresses the silent mating-type locus and subsequently leads to a diploid-like repression of the mating pathway (Dion *et al*, 2005). Since Ste12 regulates mating genes, it is possible that the repression of Ste12-bound genes in the mutant results from this effect rather than the inhibition of Ste12 binding.

As a result of this analysis, we conclude that TF binding does not depend on the acetylation state of lysines. Furthermore, as our experiments on hir3, yml081w and sum1 mutants demonstrate, the deletion of TFs does in fact affect acetylation levels. We therefore hypothesize that in most cases the TF is causing changes in acetylation and not being recruited by the acetylated lysines.

Discussion

Previous studies have attempted to determine the relationship between histone acetylation and gene expression using a

variety of techniques. For example, a recent study used regression approaches to correlate acetylation patterns with expression levels, demonstrating that the cumulative acetylation levels are predictive of expression (Yuan *et al*, 2006). An experimental effort that measured the effect of histone mutants on gene expression levels reached similar conclusions (Dion *et al*, 2005).

In contrast to these efforts, here we do not focus on the effect of acetylation on expression, but rather attempt to reconstruct the regulatory network that establishes genome-wide acetylation profiles. The enzymes that are responsible for acetylating and deacetylating histones have been largely characterized in the past couple of decades. In a limited number of cases, it is also known how they are specifically recruited to specific genomic loci to modify histone tails.

In contrast to these more limited studies, using a multiple linear regression of optimized sigmoidal function analysis, we have constructed a comprehensive and genome-wide network that links TFs and their target genes to acetylation profiles of specific lysine residues in the four core histones. For each TF, these associations are likely due to recruitment of select HATs and HDACs. We utilize protein–protein interaction data to infer which HAT or HDAC is likely recruited by each TF in the few cases where this is known. The validity of the resulting network is supported by the observation that TFs that are correlated with hypoacetylation tend to recruit deacetylases, while TFs that correlate with hyperacetylation tend to recruit acetylases. When we compare the constructed network to those that are randomly generated, we find that the enrichment of correct associations is statistically very significant.

Furthermore, we have performed experimental tests of the hypotheses generated from the network. In particular, we have measured the changes in acetylation in three deletion mutants of TFs that were present in our model: Yml081W, Hir3 and Sum1. We note a marked decrease in the acetylation of three lysines in the hir3 mutant at a histone promoter where Hir3 binds strongly, suggesting that histone acetylation may play a role in repressing the expression of histone genes. We observe increased acetylation of two lysines in two promoters where Sum1 binds, while observing little or no effect on a control lysine and a control promoter, supporting our specific predictions of which lysines and which promoters are regulated by Sum1 through the recruitment of Hst1. Finally, we have also shown that the deletion of a previously uncharacterized TF, YML081W, decreases acetylation levels at a bound promoter, as predicted by our network.

We have also shown that in general the TFs cause changes in acetylation rather than being recruited by the modifications. Beyond these observations, our model also generates more speculative predictions about the mechanisms by which the specificity of histone acetylase and deacetylase complexes is achieved. For instance, we note that the same HAT or HDAC complex appears to be recruited by different TFs through alternate subunits. It is conceivable that differential recruitment of the complex through its distinct subunits can affect its specificity for different lysines.

Our effort parallels attempts to reconstruct transcriptional regulatory networks: models of how the interactions between TFs and the promoters regulate gene expression. These efforts have advanced considerably over the past few years, providing a comprehensive network in *S. cerevisiae* (Harbison *et al*, 2004). Knowledge of this network allows us to pose questions regarding its dynamics (Luscombe *et al*, 2004) along with its evolution (Babu *et al*, 2004). We expect that as knowledge of the histone acetylation regulatory network increases and data from additional species become available, we will be able to pose analogous questions.

In future, our approach can be expanded to yield more sophisticated regulatory networks. Use of data from more precise acetylation profiles and TF binding generated using tiling arrays could reveal significant and more specific associations that would be hidden in the current data sets (Pokholok *et al*, 2005). Our methodology is also applicable to human cells, in which reconstruction of analogous networks may identify mechanisms that lead to the deregulation of histone acetylation networks in cancer cells (Seligson *et al*, 2005).

Materials and methods

Data

The histone acetylation data used in this experiment are from Kurdistani *et al* (2004). This data set includes measurements of the acetylation levels of 11 lysines found on histone tails (H2AK7, H4K8, H3K9, H2BK11, H4K12, H3K14, H2BK16, H3K18, H3K23, H3K24, H4K16) for 5366 *S. cerevisiae* promoter regions. The acetylation levels in this data set were measured relative to genomic DNA rather than nucleosomal DNA. However, since the acetylation levels of histones are affected by the occupancy of nucleosomes in that region, we divided the acetylation data by the average level of H3 and H2A histones from the Bernstein *et al* (2004) data set.

The TF-binding data set is from Harbison *et al* (2004), which includes the TF-binding levels of 203 TFs to 4956 promoters. We use the normalized and averaged ratios that are reported in Supplementary data. Only the promoter entries that are present in the acetylation, nucleosome occupancy, and TF-binding data sets were kept. In addition, since TF-binding data will serve as the basis matrix for the linear regression, the column and row entries need to be well populated for our analysis. Therefore, the rows of the acetylation data with two or more missing values are removed. In the binding data set, rows with more than 20 missing values are removed and, subsequently, columns with 50 or more missing values were also removed. After the filtering process, we are left with the binding values of 181 TFs and the acetylation levels of 11 lysines for 3221 promoters.

The other data we utilize involve protein–protein interactions between TFs and HDACs and HATs. We obtain protein–protein interactions from the *Saccharomyces* Genome Database (http://www.yeastgenome.org/). This data set was complemented with the interactions from a systematic screen of protein complexes (Gavin *et al*, 2006), the Database of Interacting Proteins (Salwinski *et al*, 2004) and the MIPS database (Salwinski *et al*, 2004) to ensure completeness. From these genome-wide data sets, we select only the interactions that contain proteins whose Gene Ontology (GO) annotation is associated with the terms 'histone acetylation/deacetylation,' 'histone acetyltransferase activity/contributes_to' and 'histone acetyltransferase/deacetylase complex' (Ashburner *et al*, 2000). The proteins that are found to be part of the histone acetylase or deacetylase complex are then linked to the HAT and HDAC protein complexes, as defined by MIPS (Salwinski *et al*, 2004).

Linear regression of optimized sigmoidal functions

Before the linear regression is performed, each column of the acetylation data is normalized to yield a mean value of zero and a variance of one. We define the TF-binding data as the basis and the

normalized acetylation data as the target. Thus, the model for the acetylation of a single lysine is

$$\log(A_{jk}) = \sum_{i=1}^{181} F_{ik}\left(\log(TF_{ji})\right)$$

where A_{jk} is the acetylation ratio for lysine k at gene j, TF_{ji} is the binding ratio of TF i to gene j and F_{ik} is the nonlinear function that describes the effect of the binding of TF i on the acetylation of lysine k. We have chosen sigmoidal functions to model this relationship (see Supplementary Figure 2),

$$F = \frac{a}{1 + e^{\frac{b-x}{c}}}$$

since these functions capture the expected saturation effects of high binding as well as the non-effects of low binding. The sigmoidal curve is a three-parameter function: the midpoint, b, the slope, c, and the maximal values, a, must all be specified. To simplify the complex optimization problem we are confronted with, we have chosen to vary only two of these parameters a and b, setting the value of c to 1 for all factors. The maximal value, a, of the sigmoidal function represents the maximal coupling strength between a TF and the acetylation of a specific lysine. If it is positive it implies that the TF is correlated with high levels of acetylation, and if it is negative it is correlated with hypoacetylation.

We use the following protocol to identify the TFs, and their functional form, that regulate each lysine. This is similar to the approach developed by Das *et al* (2006) using multiple adaptive regression splines (MARS). We first divide our genes into two groups: test and training. The test genes are 10% of the total and the training genes are the remaining 90%. Using the training genes, we greedily build up a model containing successively more TFs. We use the reduction in variance of our model as the figure of merit:

$$RIV = 1 - \frac{\text{var(residual)}}{\text{var(data)}}$$

where the residual is the difference between the observed and modeled data. After we select the best factor, we select the second that in combination with the first maximally reduces the variance of the residual. We then iterate this procedure until we reach a maximum model size of 25. At each step, we monitor the reduction in variance of the test set and consider the model size with the lowest variance as the optimum. We repeat all these steps ten times for 10 different randomly selected test sets.

This approach generates 10 independent models. From these, we generate a single consensus model, by selecting the top TFs that appear in all 10 models, making sure that the consensus model is not larger than the average optimal model size. Finally, we compute the reduction in variance of this final model on a randomly selected test data set, as shown in Supplementary Figure 4. We see that on average these models have a reduction in variance of about 0.15. The full model is described in Supplementary Table 1, including all the factors and their parameters for each lysine. We also report the computed changes in acetylation that arise from the deletion of TFs at target promoters in Supplementary Table 2.

Linear regression of optimized sigmoidal function models of expression profiles is computed using the same protocol, simply replacing the acetylation profile with the logarithm of the expression ratios.

Statistical validation of network

The validation test is built on the premise that TFs that are correlated with hyperacetylation should recruit HATs while factors that are correlated with hypoacetylation should recruit HDACs. In practice, we test for the number of associations in our network between factors with positive alpha coefficients and HATs and negative alpha coefficients and HDACs. The statistical significance of this number is computed using the hypergeometric cumulative distribution, which computes the probability of correct matches, or greater, given the number of positive and negative alpha coefficient TFs and the number of HATs and HDACs in our network.

Network construction

From the linear regression computation described above, we define the TFs that regulate each lysine. We use these couplings to form the edges between the TFs and lysines. To form the link between TFs and HATs and HDACs, we use protein–protein interactions from the integrated data set described above.

The TFs are displayed in various shades of blue and red. The blue nodes indicate TFs that are correlated with histone acetylation, the red nodes indicate TFs that are linked to histone deacetylation and white suggests neither activity. The various shades of blue and red for each TF were derived from the maximal value of their corresponding sigmoidal function. The strength of the color is determined by the summing of maximal values associated with the same TF across all the lysines it is associated with. The TF with the darkest shade of blue has the maximum positive sum and the TF with the darkest shade of red has the most negative sum, with all other shades of red and blue being in between. Finally, the histone acetylases, histone deacetylases and protein complexes in the third and topmost tiers are colored blue and red as well to indicate histone acetyltransferase and deacetylase activity, respectively, as defined by GO and MIPS.

The individual links between the TFs and lysines follow the same coloring scheme as above. In this case, the color intensity of the link is determined by the maximal value of the sigmoidal function between the TF and the lysine.

Chromatin immunoprecipitation

Chromatin immunoprecipitation has been performed as previously described (Suka *et al*, 2001). Briefly, yeast strains were grown to OD_{600} of 0.8 in YPD media. DNA was fragmented with an average size of approximately 500 bp by sonication and 50 µl of the lysate was immunoprecipitated using 0.5 µl of H3 antibody (Abcam), 2 µl of H3K18Ac-specific antibody and 4 µl of H4K16-specific antibody. PCR was performed within the linear range of amplification for each primer set and DNA sample. Typically, 1/100th of the immunoprecipitated DNA and 0.8 µCi/µl [α-^{32}P]dATP (specific activity, 3000 Ci/mmol) were used for each PCR (12.5 µl). Quantitations of ChIP data were performed using a PhosphorImager and ImageQuant software (Molecular Dynamics, Sunnyvale, CA). All PCR reactions include primers designed to a region 500 bp from the telomere of chromosome VI-R as an internal control for loading and quantitation.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

References

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat Genet* **25:** 25–29

Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA (2004) Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol* **14:** 283–291

Bernstein BE, Liu CL, Humphrey EL, Perlstein EO, Schreiber SL (2004) Global nucleosome occupancy in yeast. *Genome Biol* **5:** R62

Das D, Nahle Z, Zhang MQ (2006) Adaptively inferring human transcriptional subnetworks. *Mol Syst Biol* **2:** 2006 0029

Dion MF, Altschuler SJ, Wu LF, Rando OJ (2005) Genomic characterization reveals a simple histone H4 acetylation code. *Proc Natl Acad Sci U S A* **102:** 5501–5506

Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B, Edelmann A, Heurtier MA, Hoffman V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick JM, Kuster B, Bork P, Russell RB, Superti-Furga G (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440:** 631–636

Guo X, Tatsuoka K, Liu R (2006) Histone acetylation and transcriptional regulation in the genome of *Saccharomyces cerevisiae*. *Bioinformatics* **22:** 392–399

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* **431:** 99–104

Hu Z, Killion PJ, Iyer VR (2007) Genetic reconstruction of a functional transcriptional regulatory network. *Nat Genet* **39:** 683–687

Jenuwein T, Allis CD (2001) Translating the histone code. *Science* **293:** 1074–1080

Kurdistani SK, Grunstein M (2003) Histone acetylation and deacetylation in yeast. *Nat Rev Mol Cell Biol* **4:** 276–284

Kurdistani SK, Tavazoie S, Grunstein M (2004) Mapping global histone acetylation patterns to gene expression. *Cell* **117:** 721–733

Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431:** 308–312

Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E, Zeitlinger J, Lewitter F, Gifford DK, Young RA (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122:** 517–527

Prochasson P, Florens L, Swanson SK, Washburn MP, Workman JL (2005) The HIR corepressor complex binds to nucleosomes generating a distinct protein/DNA complex resistant to remodeling by SWI/SNF. *Genes Dev* **19:** 2534–2539

Richmond TJ, Davey CA (2003) The structure of DNA in the nucleosome core. *Nature* **423:** 145–150

Rundlett SE, Carmen AA, Suka N, Turner BM, Grunstein M (1998) Transcriptional repression by UME6 involves deacetylation of lysine 5 of histone H4 by RPD3. *Nature* **392:** 831–835

Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D (2004) The database of interacting proteins: 2004 update. *Nucleic Acids Res* **32:** D449–D451

Seligson DB, Horvath S, Shi T, Yu H, Tze S, Grunstein M, Kurdistani SK (2005) Global histone modification patterns predict risk of prostate cancer recurrence. *Nature* **435:** 1262–1266

Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* **9:** 3273–3297

Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* **403:** 41–45

Suka N, Suka Y, Carmen AA, Wu J, Grunstein M (2001) Highly specific antibodies determine histone acetylation site usage in yeast heterochromatin and euchromatin. *Mol Cell* **8:** 473–479

Xie J, Pierce M, Gailus-Durner V, Wagner M, Winter E, Vershon AK (1999) Sum1 and Hst1 repress middle sporulation-specific gene expression during mitosis in *Saccharomyces cerevisiae*. *EMBO J* **18:** 6448–6454

Xu F, Zhang K, Grunstein M (2005) Acetylation in histone H3 globular domain regulates gene expression in yeast. *Cell* **121:** 375–385

Yuan GC, Ma P, Zhong W, Liu JS (2006) Statistical assessment of the global regulatory role of histone acetylation in *Saccharomyces cerevisiae*. *Genome Biol* **7:** R70