Cell, Vol. 117, 721–733, June 11, 2004, Copyright ©2004 by Cell Press

Mapping Global Histone Acetylation Patterns to Gene Expression

Siavash K. Kurdistani,[1] Saeed Tavazoie,[2,*] and Michael Grunstein[1,*]

[1]Department of Biological Chemistry
UCLA School of Medicine and the Molecular Biology Institute
Boyer Hall
Los Angeles, California 90095

[2]Department of Molecular Biology
and The Lewis-Sigler Institute for Integrative Genomics
Princeton University
Princeton, New Jersey 08544

Summary

Histone acetyltransferases and deacetylases with specificities for different sites of acetylation affect common chromatin regions. This could generate unique patterns of acetylation that may specify downstream biological processes. To search for existence of these patterns and their relationship to gene activity, we analyzed the genome-wide acetylation profiles for eleven lysines in the four core histones of *Saccharomyces cerevisiae*. We find that both hyper- and hypoacetylation of individual lysines are associated with transcription, generating distinct patterns of acetylation that define groups of biologically related genes. The genes within these groups are significantly coexpressed, mediate similar physiological processes, share unique *cis*-regulatory DNA motifs, and are enriched for binding of specific transcription factors. Our data also indicate that the in vivo binding of the transcription factor Bdf1 is associated with acetylation on most lysines but relative deacetylation on H4 lysine 16. Thus, certain acetylation patterns may be used as surfaces for specific protein-histone interactions, providing one mechanism for coordinate regulation of chromatin processes that are biologically related.

Introduction

Acetylation and deacetylation of lysine residues in the N termini of histones play fundamental roles in diverse chromatin-based processes. Such roles have been established through genetic and biochemical analyses of individual sites of acetylation as well as the discovery of histone acetyltransferases (HAT) and deacetylases (HDAC) as coregulators in transcription, DNA replication, and DNA repair (Kurdistani and Grunstein, 2003). The sites of acetylation include at least four highly conserved lysines in histone H4 (K5, K8, K12 and K16), five in histone H3 (K9, K14, K18, K23 and K27), as well as less conserved sites in histones H2A and H2B. Various HATs and HDACs show distinct specificities toward these acetylation sites in yeast. For instance, the Gcn5 HAT is required for acetylation of H3 and H2B sites at the *INO1* and *IME2* promoters in vivo. In contrast, inactivation of Esa1 HAT activity in vivo results in loss of acetylation mainly on H4 and H2A at these same promoters (Suka et al., 2001). In the case of HDACs, yeast Rpd3 is required for deacetylation of all core histones (Suka et al., 2001), while Hda1 and Hos2 specifically deacetylate histones H3/H2B (Wu et al., 2001) and H4/H3 (Wang et al., 2002), respectively.

The mechanisms by which histone acetylation and deacetylation affect transcription (and other DNA-based processes) are thought to involve two major pathways. First, histone acetylation may alter the folding properties of the chromatin fiber, thereby modulating the accessibility of DNA through structural changes (Horn and Peterson, 2002; Tse et al., 1998). Second, the lysine residues and their modifications also provide specific binding surfaces for the recruitment of repressors and activators of gene activity. In the formation of yeast heterochromatin, for example, the silencing information regulator Sir3 interacts with a domain of histone H4 (residues 16-29) that contains the acetylatable lysine 16 (H4K16). Genetic, chromatin immunoprecipitation (ChIP) and in vitro binding data argue that H4K16 must be deacetylated for the Sir3-H4 interaction to occur (Kurdistani and Grunstein, 2003). Similarly, the ISWI protein, a member of a nucleosome-remodeling complex associated with inactive chromatin in flies, also interacts with the domain of H4 that includes K16 which, when acetylated, disrupts the ISWI-H4 interaction (Corona et al., 2002; Deuring et al., 2000). Hence, the acetylation state of a single lysine residue can regulate the binding of proteins to chromatin.

In euchromatin, the relationship between specific sites of acetylation and transcription has been less clear. In *Saccharomyces cerevisiae*, mutagenesis of acetylation sites in histone H4 (lysines 5, 8, 12, and 16) decreases activation of the *GAL1, GAL7*, and *GAL10* genes whereas similar mutagenesis of lysine residues at positions 9, 14, 18, and 23 in histone H3 results in hyperactivation of these genes (Mann and Grunstein, 1992). Selective acetylation of H4 K5/K12 occurs in the cytoplasm prior to H4 assembly in chromatin (Sobel et al., 1995). Although this K5/K12 "deposition pattern" is not required for assembly per se, it may be associated with gene activity (Ma et al., 1998). Yet other findings point to the general use of H4K8/K16 acetylation in transcription (Strahl and Allis, 2000). Studies utilizing mammalian cells identify H4K8 and H3K14 as required acetylation sites for recruitment of transcription-related complexes for activation specifically of the human *IFN*-β gene (Agalioti et al., 2002). These data indicate that acetylation of different lysine residues have distinct roles in gene activity at the level of single genes and/or the whole genome. These findings have led to the proposal that histone acetylation (and other modifications) act as "codes" for effecting downstream events (Strahl and Allis, 2000). However, the elements of a code must be both consistent and combinatorial for the code to exist (Kurdistani and Grunstein, 2003). To date, there is no evidence for

*Correspondence: mg@mbi.ucla.edu (M.G.), tavazoie@molbio.princeton.edu (S.T.)

consistent patterns of acetylation from gene to gene or for the combinatorial use of histone modification sites. As a result, an alternative hypothesis argues that histone modifications do not constitute a code but are part of the process of protein signaling (Schreiber and Bernstein, 2002).

In trying to decipher the role of acetylation sites as signals for transcription, it is important to understand the binding properties of proteins that interact with histones. While Sir3 and ISWI bind to the unacetylated H4K16 in heterochromatin or repressed chromatin, the bromodomain protein module has been shown to bind acetylated lysines in vitro (Dhalluin et al., 1999; Jacobson et al., 2000). Bromodomains are found in several eukaryotic transcription factors including Gcn5 and Taf1 (TFII250), which is the largest subunit of the TFIID transcription initiation complex, as well as in proteins in chromatin-remodeling complexes such as Snf2 of the yeast SWI/SNF complex. Metazoan Taf1, which contains two bromodomains, binds the diacetylated H4 tail (K5/K12 or K8/K16) better than mono- or unacetylated tails in vitro (Jacobson et al., 2000). The yeast Taf1 ortholog lacks bromodomains but instead may use the bromodomain factor 1 (Bdf1). In vitro, Bdf1 preferentially binds tetraacetylated histone H4 (K5/8/12/16) and hyperacetylated histone H3, but not acetylated H2A and H2B or unacetylated H4 and H3 (Ladurner et al., 2003; Matangkasombut and Buratowski, 2003). Considering that the in vitro specificity of Bdf1 has been examined through its interaction with short peptides, single histone proteins, or histone tetramers, it is important to determine the in vivo preference of Bdf1 for individual acetylation sites in native chromatin.

The experimental approaches toward understanding the biology of histone acetylation have so far been by and large reductionist, focusing on one or two lysines or examination of a few genomic loci at a time. The recent analysis of genome-wide HDAC activity has addressed the chromosomal sites of action for five related deacetylases when they are deleted, but not the relationship between acetylation and transcription when deacetylases are functional (Robyr et al., 2002). The spatial and functional heterogeneity of histone acetylation sites and the proteins that bind to them require that we ascend in scale to genome-wide views to discern patterns that are less apparent in gene-by-gene surveys. Here, we utilized the DNA microarrays for both intergenic regions (IGRs) and open reading frames (ORFs) to determine the genome-wide patterns of 11 sites of acetylation on the four core histones (H4K8, 12, 16; H3K9, 14, 18, 23, 27; H2AK7; H2BK11, 16). We apply a systematic set of statistical and clustering algorithms to explore the relationship between these sites and their correlation with gene activity. We determine that combinations of histone acetylation sites on IGRs or, remarkably on ORFs, can generate consistent patterns of acetylation for biologically related genes. To explore the role of these patterns, we identify the preference of Bdf1 for individual lysines in vivo. These findings suggest that unique histone acetylation patterns, involving combinations of hyper- and hypoacetylated lysines, may provide binding surfaces for certain regulatory factors to coordinate the activity of related groups of genes.

Results

Acetylation DNA Microarrays and Data Collection

Acetylation microarrays combine chromatin immunoprecipitation (ChIP) with highly specific antibodies against individual sites of acetylation and DNA microarrays to determine the in vivo levels of histone acetylation at various loci throughout the genome (Robyr et al., 2002; Suka et al., 2001). In this study, chromatin fragments were isolated from exponentially growing YDS2 cells, treated with formaldehyde and immunoprecipitated (IP) with acetylation site-specific antibodies. These antibodies were shown previously to be highly specific in ChIP experiments and do not crossreact significantly with other acetylation sites (Suka et al., 2001). DNA from input and IP fractions were purified, amplified by PCR, and labeled with either Cy3 or Cy5 fluorophores. The labeled DNA from input and IP fractions were combined and hybridized to a microarray glass slide containing about 6700 IGRs (prepared in house) or more than 6200 ORFs (University Health Networks, Toronto). All experiments were repeated 2–4 times from samples separately crosslinked for ChIP. At least one set of experiments was carried out with the fluorescent dyes switched between the input and IP samples to control for the efficiency of labeling. Each set of ChIP experiments by the 11 antibodies for both IGRs and ORFs was done on the same starting extract. Array data were normalized by the ratio of total intensities over the entire array between the two fluorescent dyes (IP/Input) and averaged across replicate experiments. To increase reliability of the data, only data points with a coefficient of variation less than 0.5 ($CV < 0.5$) between replicate experiments were considered. Shared intergenic regions between divergent ORFs were assigned to both. All analyses were done for genomic loci for which data on all 11 acetylation sites were available (2206 IGRs and 2403 ORFs). In addition to comparing the level of acetylation of each lysine as obtained from the arrays, we variance-normalized the acetylation profile of each IGR and ORF across the 11 sites (see Experimental Procedures). Variance-normalization reflects the relative acetylation level of each site compared to all other sites. IGR versus ORF acetylation are denoted by "i" and "o" prefixes, respectively.

Correlation between Individual Sites of Acetylation

We first wished to compare the acetylation level of each lysine to all other lysines within IGRs and within ORFs. The data are presented graphically for the acetylation levels without variance-normalization (Figures 1A and 1B) and after variance-normalization (Figures 1D and 1E) as color-coded correlation matrices according to the scales shown. When we compare the acetylation levels of each lysine to those of other lysines for IGRs, we find strong positive correlations between certain acetylation sites on the same or different histones (quantitations shown also as bar graphs in Figure 1C). For example, H4 iK8 and iK12 are highly correlated with each other ($r = 0.83$, $P < 10^{-22}$), as are H4 iK8 and H3 iK14 ($r = 0.78$, $P < 10^{-22}$) and H2B iK11 and iK16 sites ($r = 0.88$, $P < 10^{-22}$). However, we find significantly weaker correlations between H4 iK16 and H3 iK9/18/27, especially with iK18 ($r = 0.35$, $P < 10^{-22}$).

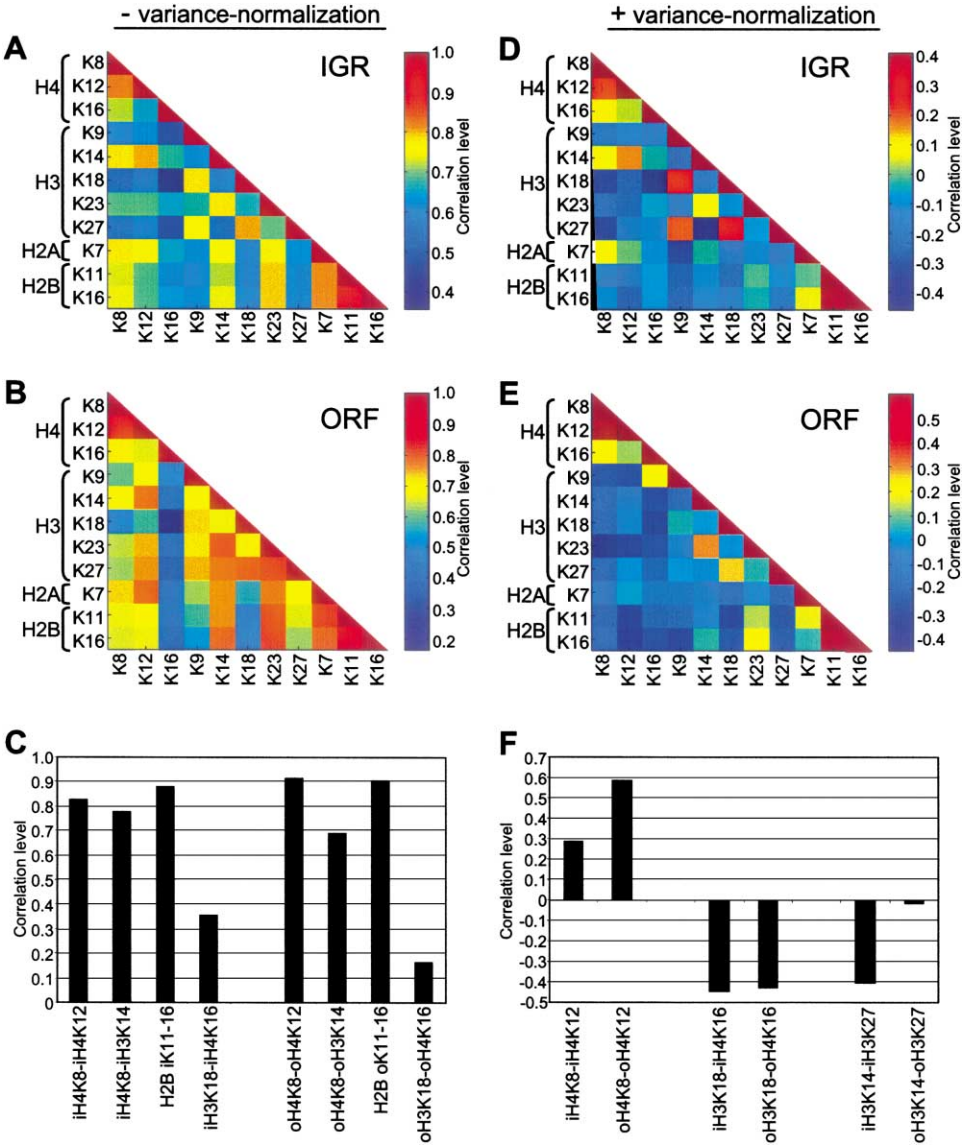

Figure 1. Acetylation Levels of Lysines Are Positively and Negatively Correlated to Each Other

(A–B) Matrices depicting the degree of correlation between acetylation levels of different sites as obtained form the (A) intergenic (IGR, "i") and (B) open-reading frame (ORF, "o") arrays without variance-normalization. The color scale indicates the degree of correlation. The diagonal is the correlation of each site to itself.

(C) Correlation values for the acetylation levels of indicated lysine pairs are depicted as bar graphs.

(D–E) Matrices depicting the degree of correlation between acetylation levels of different sites after variance normalization of the acetylation profile of each (D) IGR and (E) ORF across the 11 sites. These values reflect the relative acetylation level of each site compared to all other sites. The diagonal is the correlation of each site to itself. Note that darker shades of blue indicate increasing anticorrelation. A high degree of correlation (darker shades of red) between two lysines suggests that in relation to other sites of acetylation, the lysine pair tend to be acetylated or deacetylated to the same extent. A high degree of anticorrelation (darker shades of blue) between two lysines suggests that in relation to other sites of acetylation, the lysine pair tend to be acetylated or deacetylated oppositely.

(F) Correlation values for the variance-normalized acetylation levels of indicated lysine pairs are depicted as bar graphs.

Acetylation levels on the ORFs show similar levels of correlations as those between sites on the IGRs (Figures 1B and 1C). For example, H4 oK8 and oK12 show a strong positive correlation to each other ($r = 0.91$, $P < 10^{-22}$), as do H4 oK8 and H3 oK14 ($r = 0.70$, $P < 10^{-22}$) and H2B oK11 and oK16 ($r = 0.91$, $P < 10^{-22}$). H4 oK16 shows the weakest correlation with H3 oK9/18/27 acetylation sites, especially with oK18 ($r = 0.17$, $P < 10^{-7}$). Thus, at a given genomic locus (IGR or ORF), individual lysines are acetylated to varying extents, the least correlated of which are H4K16 and H3K18.

Since IGR and ORF data were obtained from separate arrays, the acetylation values between these regions were not compared. To perform such analyses, it is important to obtain acetylation data from arrays that contain both intergenic and coding regions on the same slide.

In comparing the relative acetylation levels (plus vari-

Table 1. Correlation (*r*) of Sites of Acetylation on Intergenic Regions with Gene Expression

Lysine	r	P values
iH4 K8	−0.15	1.50E-12
iH4 K12	−0.08	1.69E-05
iH4 K16	−0.18	4.14E-16
iH3 K9	0.07	1.08E-03
iH3 K14	−0.10	7.54E-06
iH3 K18	0.21	4.68E-23
iH3 K23	−0.04	5.80E-02
iH3 K27	0.08	2.27E-04
iH2A K7	−0.10	1.19E-05
iH2B K11	−0.13	5.08E-09
iH2B K16	−0.13	6.67E-09

i, intergenic regions; o, open reading frames

Table 2. Correlation (*r*) of Sites of Acetylation on Coding Regions with Gene Expression

Lysine	r	P values
oH4 K8	−0.03	8.51E-02
oH4 K12	0.03	1.73E-01
oH4 K16	−0.12	1.86E-08
oH3 K9	0.10	1.20E-06
oH3 K14	0.05	3.31E-02
oH3 K18	0.36	4.28E-23
oH3 K23	0.07	4.89E-04
oH3 K27	0.20	5.03E-23
oH2A K7	−0.02	3.28E-01
oH2B K11	−0.14	9.18E-12
oH2B K16	−0.12	3.14E-09

i, intergenic regions; o, open reading frames

ance-normalization), we find that sites of acetylation are both positively and negatively correlated with each other (Figures 1D–1F). For instance, H4K8 and K12 are positively correlated with each other on IGRs ($r = 0.28$, $P < 10^{-19}$) and on ORFs ($r = 0.59$, $P < 10^{-22}$). Such positive correlations indicate that, in relation to other sites of acetylation, K8 and K12 tend to be acetylated to the same extent at a genomic locus. In contrast, H4K16 and H3K18 are anticorrelated at $r = -0.44$ for IGRs ($P < 10^{-22}$) and $r = -0.43$ for ORFs ($P < 10^{-22}$). This means that for a genomic region, if K18 acetylation is higher than average acetylation for that region, then it is likely that the level of K16 acetylation is lower than average and vice versa. Variance-normalized relative acetylation levels also show significant differences within IGRs and ORFs. For instance, H3 iK14 and iK27 are strongly anticorrelated ($r = -0.43$; $P < 10^{-22}$) with each other but oK14 and oK27 show virtually no correlation ($r = -0.02$), indicating that K14 and K27 acetylation levels are inversely regulated on IGRs but not on ORFs. Altogether, these data indicate that acetylation levels of individual lysines may be regulated in relation to acetylation levels of other lysines in the same or different histones. Below, we show that these differences are important for defining groups of coregulated genes.

Both Hyper- and Hypoacetylation of Individual Sites Are Correlated with Gene Activity

Although histone acetylation and deacetylation have been linked to transcriptional activation and repression, respectively, the relation of each acetylation site to transcription has not been determined on a genome-wide basis. Thus, we determined such global correlations between individual acetylation sites and gene activity (Causton et al., 2001) (Tables 1 and 2). Comparing acetylation of each lysine on IGRs with the expression of its immediate downstream gene (Table 1), we find that H3 iK18 acetylation is best correlated with high transcriptional activity ($r = 0.21$, $P < 4.7 \times 10^{-23}$) followed by acetylation of H3 iK27 ($r = 0.08$, $P < 2.3 \times 10^{-4}$) and iK9 ($r = 0.07$, $P < 1.1 \times 10^{-3}$). To our surprise, acetylation of most other sites are negatively correlated with increased transcription. Most significant is the anticorrelation of H4 iK16 acetylation with transcription ($r = -0.18$, $P < 4.1 \times 10^{-16}$) followed by other H4 sites iK8 ($r = -0.15$, $P < 1.5 \times 10^{-12}$) and iK12 ($r = -0.09$, $P < 1.7 \times 10^{-5}$). Both histones H2A and H2B acetylation sites are also negatively correlated with transcriptional activity. It should be noted that these correlations, while highly significant, may be artificially low due to technical limitations. For example, the DNA sequences on the arrays are on average 1 kb in length and may contain several nucleosomes with different acetylation levels.

Acetylation of lysines on ORFs shows a similar pattern of correlations with gene activity with some exceptions (Table 2). As in the case of IGRs, H3 oK18, oK27, and oK9 acetylation are best correlated with high transcriptional activity in that order. Acetylation of H4 oK16 is also negatively correlated with transcriptional activity but other H4 sites, oK8 and oK12 show no correlation. While H2A oK7 acetylation is not related to gene activity, the H2B acetylation sites, oK11 and oK16 show the most significant negative correlations with transcription.

Therefore, on both intergenic and coding regions, hyperacetylation of histone H3K9/18/27 but hypoacetylation of H4K16 and H2BK11/16 are correlated with transcription. These results indicate that both acetylation and deacetylation of individual lysine residues may be involved in transcriptional regulation.

Systematic Identification of Clusters with Complex Acetylation Patterns

The correlations described above represent the relationship between gene expression and an individual site of acetylation over the entire genome. To determine whether unique patterns of histone acetylation, involving combinations of the 11 sites, are associated with distinct groups of genes, we applied the *k*-means clustering algorithm (Hartigan, 1975) to the variance-normalized acetylation profile of each IGR or ORF. *k*-means is a partitional clustering method that by iterative reallocation of candidate members defines cohesive groups of genes based on pattern similarity. We constrained the algorithm to identify clusters of at least 20 genes with a correlation $C \geq 0.7$ as a measure of similarity among the members' acetylation profiles for all 11 sites.

For IGRs, we identified 53 clusters of 22–62 genes per cluster, comprising 1771 genes. Figure 2A depicts the average acetylation levels for members of each IGR cluster in a pictorial format. The acetylation levels are color-coded according to the scale shown. Each row

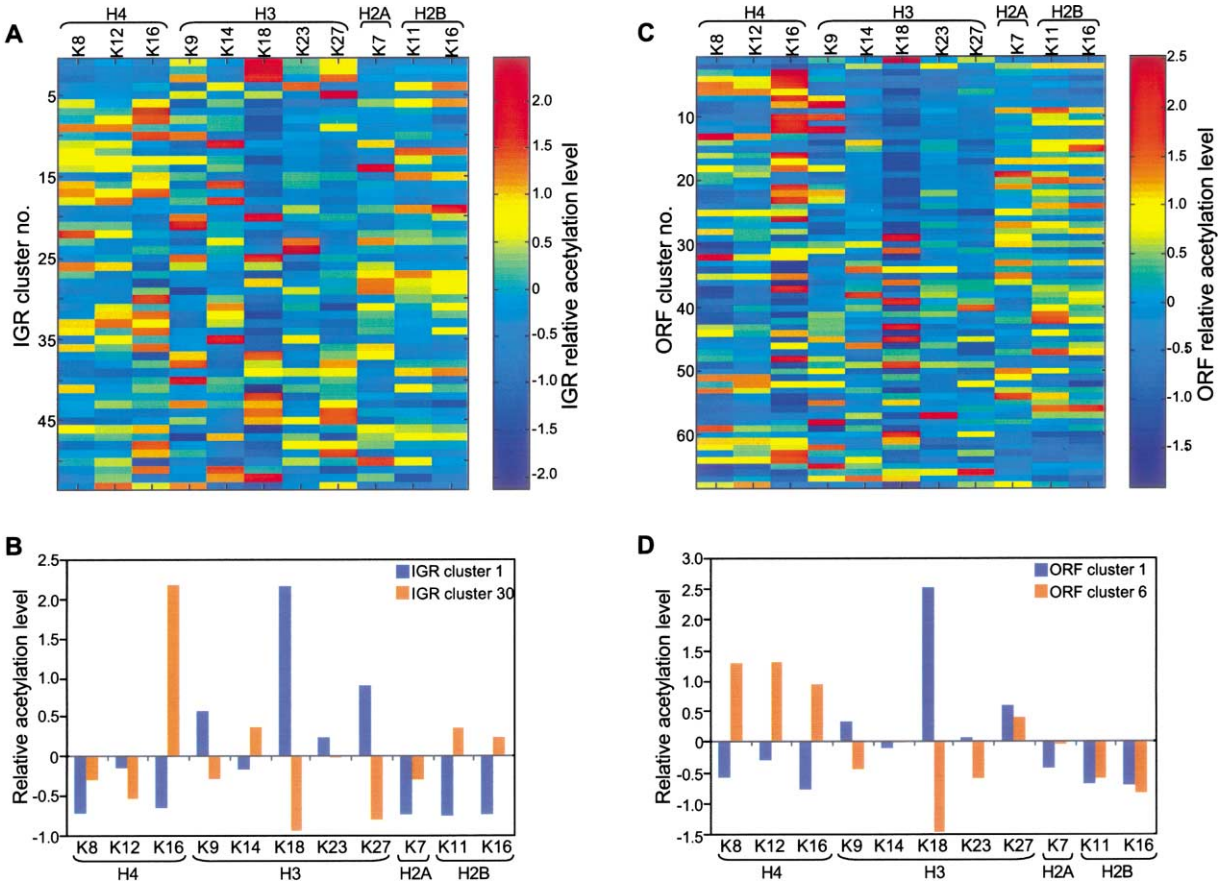

Figure 2. Groups of Genes Share Similar Acetylation Patterns on Their Promoter or Coding Regions

(A) Average acetylation patterns of genes in the IGR clusters. Each row represents one cluster with average acetylation of each lysine in columns as marked and color-coded according to the scale shown. Darker shades of red indicate increased acetylation; darker shades of blue indicate decreased acetylation. The correlation between patterns of acetylation for genes in any cluster is 0.7 or higher.

(B) Average acetylation patterns of genes in IGR clusters 1 and 30 as bar graphs.

(C) Average acetylation patterns of genes in the ORF clusters presented as in (A).

(D) Average acetylation patterns of genes in ORF clusters 1 and 6 as bar graphs.

represents one cluster with average acetylation values of individual lysines corresponding to the columns as marked (K8, K12, K16, etc.). For instance, in Figure 2A, rows (clusters) 1, 2, and 3 all contain genes with a relatively high level of acetylation at H3K9/18/27 and relatively low levels at other lysines in their upstream regions. In contrast, the genes of row (cluster) 30, on average contain relatively hyperacetylated H4K16 as compared to the other lysines. The average acetylation values for clusters 1 and 30 are also shown as bar graphs in Figure 2B. Other clusters also show very different patterns. While cluster 21 is predominantly acetylated on H3K9, in cluster 29, the same lysine (K9) is deacetylated but H2AK7 is mainly acetylated.

We were surprised to find that the ORFs also partition into clusters based solely on the coding region patterns of acetylation and independent of any knowledge about their upstream IGR acetylation profiles. For ORFs, we identified 68 clusters of 22–56 genes per cluster, comprising 2253 coding regions (Figure 2C). ORF clusters also show complex patterns of acetylation involving different acetylation sites. For instance, ORF cluster 1 is primarily acetylated on H3K18 with relative hypoacetylation of H4 sites (Figure 2D). ORF cluster 6 shows the opposite pattern with H4 sites (K8/12/16) being hyper- and H3K18 being hypoacetylated. Interestingly, genes in IGR and ORF clusters 1 are similarly acetylated on their promoter and coding regions (compare Figures 2B and 2D). As indicated below, these clusters are primarily comprised of highly expressed genes involved in anabolic processes (e.g., ribosomal protein genes), consistent with the positive correlation between H3K18 acetylation and gene activity. Therefore, we conclude that complex but unique patterns of acetylation in promoter or coding regions are common to distinct groups of genes.

Acetylation Clusters Are Enriched for Coexpressed Genes

To determine whether these acetylation clusters are biologically relevant, we asked whether the IGR and ORF acetylation clusters are each enriched for genes that have similar expression levels in exponentially growing cells, the same condition in which acetylation data were

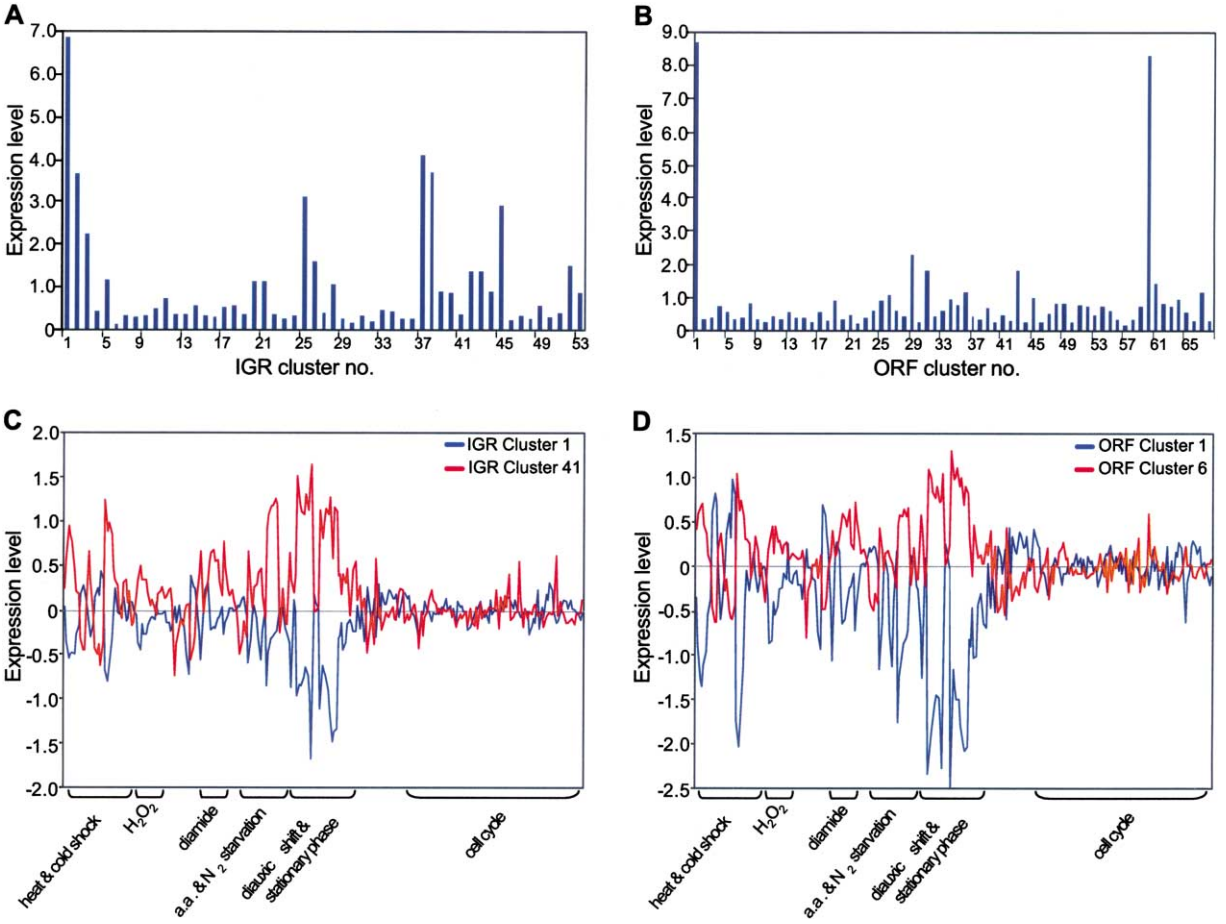

Figure 3. Genes in the Acetylation Clusters Are Coexpressed

(A) Average expression levels of genes during exponential growth in each IGR cluster. Clusters 1, 2, 6, 8, 23, 25, 29, 30, 32, 35, 36, 37, 45, and 46 have significant high or low expression at $P < 0.01$.

(B) Average expression levels of genes during exponential growth in each ORF cluster. Clusters 1, 2, 10, 16, 18, 22, 30, 38, 42, 44, 46, 50, 60, and 66 have significant high or low expression at $P < 0.01$.

(C) The genes in IGR clusters 1 (blue) and 41 (red) show significant coexpression across 255 conditions at $P < 0.001$. Some of these conditions are indicated as marked (for a complete list, see Supplemental Data available on *Cell* website). The expression patterns of these clusters are anti-correlated ($r = -0.77$)

(D) The genes in ORF clusters 1 (blue) and 6 (red) show significant coexpression across 255 conditions at $P < 0.001$ but are anticorrelated ($r = -0.84$) with respect to each other.

obtained. If so, the average expression level of the genes within each cluster would be expected to differ significantly (over- or underexpressed) from a randomly selected gene group of similar size. Figure 3 depicts the average expression levels for each IGR (Figure 3A) and ORF (Figure 3B) cluster. We found 14/53 IGR and 15/68 ORF clusters to have significantly high or low expression levels at $P < 0.01$. For instance, the genes in IGR and ORF clusters 1 are highly expressed whereas those of IGR cluster 6 and ORF cluster 30 are highly repressed. Therefore, acetylation clusters define groups of genes with similar levels of transcriptional activity.

To establish whether the genes in acetylation clusters are also coexpressed in other conditions, we determined the clusters' average expression levels across 255 conditions which were compiled from environmental stresses including heat and cold shock, amino acid (aa) and nitrogen (N_2) starvation, diauxic shift and stationary phase, and cell cycle expression data (Beer and Tavazoie, 1999). For a given cluster, we measured the correlation between expression patterns of each pair of genes for all gene pairs in that cluster and calculated *P*-values for observing an extreme median correlation coefficient (Figure S1). For IGRs, we found significant coexpression in clusters 1 and 41 at $P < 0.001$, clusters 9 and 17 at $P < 0.01$ and clusters 7 and 53 at $P < 0.05$. For instance, as graphed in Figure 3C, genes in cluster 1 are downregulated under aa/N_2 starvation, diauxic transition, and stationary phase but genes in cluster 41 are upregulated in the same conditions. Interestingly, both the acetylation profiles and the expression patterns of these clusters are strongly anticorrelated with a correlation coeffi-

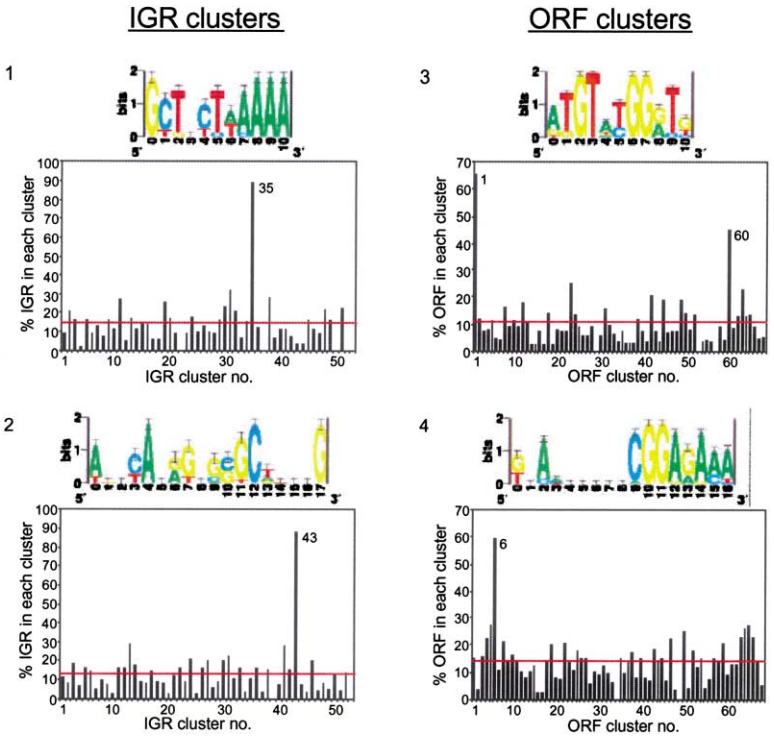

Figure 4. Genes in the Acetylation Clusters Share Common DNA Sequence Motifs

We identified upstream *cis*-regulatory sequences that are common to the genes within IGR and ORF clusters at $P < 0.0001$. Occurrences of two motifs each for IGR (right panel) and ORF (left panel) clusters are shown along with the consensus sequence for the found motif in sequence logo representation. The overall height of the stack reflects positional information content of the sequence (0–2 bits). The height of each letter is proportional to its frequency, with the most frequent on top. Certain motifs have known biological roles such as motif 3 in ORF cluster 1 which is similar to the Rap1 binding element (the same Rap1 motif is found upstream of genes in ORF cluster 60 but is not significant at $P < 0.0001$). The dotted red line is the overall genomic prevalence of the identified motifs.

cient $r = -0.74$ for acetylation (Figure 2A) and $r = -0.77$ for expression. Thus IGR clusters consist of genes that are coexpressed under multiple conditions.

For ORFs, we found significant coexpression in clusters 1, 6, 29, and 60 at $P < 0.001$; clusters 40, 54, 61, and 68 at $P < 0.01$; and clusters 13, 16, 18, 29, and 57 at $P < 0.05$. The expression patterns of ORF clusters 1 and 6 are presented in Figure 3D. These clusters show significant but opposite patterns of gene expression ($r = -0.84$). Interestingly, acetylation patterns of these clusters are also anticorrelated ($r = -0.52$, Figure 2D). From these data, we conclude that genes within IGR and ORF acetylation clusters are coexpressed under a range of conditions and, therefore, are biologically related.

Acetylation Clusters Are Enriched for Genes of Specific Functional Categories

Since members of acetylation clusters are coexpressed, we asked whether they share other common biological characteristics such as similar functions, subcellular localization, protein complexes, etc. We mapped each acetylation cluster to several classification databases including MIPS, GO, MDS, Cellzome Proteomics Complexes, and Proteome Localization using FunSpec (see Experimental Procedures). Statistical analyses revealed that 12 of 53 IGR clusters and 13 of 68 ORF clusters were enriched significantly for genes involved in common biological processes. As listed in Supplemental Table S1 available at http://www.cell.com/cgi/content/full/117/6/721/DC1, these processes include Protein Synthesis (IGR clusters 1 and 2; ORF clusters 1 and 60; $P < 10^{-12}$), endoplasmic reticulum and nuclear envelope localization (IGR cluster 37, $P < 10^{-5}$), C-compound and carbohydrate metabolism (ORF cluster 6, $P < 10^{-6}$) and mitochondrial localization (ORF cluster 14, $P < 10^{-4}$). Together with the expression coherence, functional category enrichments indicate strongly that the acetylation clusters define groups of biologically related genes.

Acetylation Clusters Contain Unique DNA Sequence Motifs

Since sequence-specific transcription factors may recruit HATs and HDACs to particular promoters, *cis*-regulatory elements that are specific to an acetylation cluster could potentially contribute to the establishment of similar acetylation patterns and eventually to coregulation. Using the AlignACE algorithm (Hughes et al., 2000), we sought to identify regulatory motifs within IGRs or in the DNA sequences upstream of ORFs. We found that 102 motifs from 29 of 53 IGR clusters, and 110 motifs from 34 of 68 ORF clusters passed our criteria for enrichment at $P < 0.0001$ (see Experimental Procedures). The occurrence and consensus sequence of representative motifs from two IGR and two ORF clusters are shown in Figure 4. Although most identified motifs have not been previously characterized (motifs 1, 2, and 4 in Figure 4), a few are binding elements for known transcriptional regulators. For instance, an identified motif in ORF cluster 1 (motif 3 in Figure 4) is highly similar to the Rap1 binding site that is found upstream of many ribosomal protein genes, which are enriched in this cluster (Supplemental Table S1 available on *Cell* website). A similar Rap1 motif is identified in IGR cluster 1. Interestingly, in addition to the uncharacterized motif shown in Figure 4 (motif 4), ORF cluster 6 is enriched for the STRE (stress-

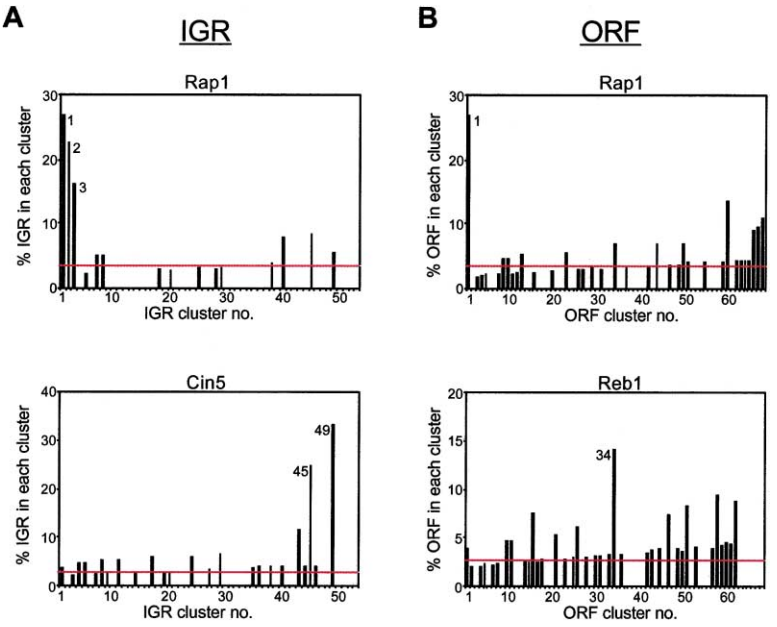

Figure 5. Genes in Acetylation Clusters Are Enriched for Specific Transcription Factor Binding

Mapping of binding data for 25 transcription factors to the acetylation clusters revealed significant enrichment in 14 IGR and 4 ORF clusters at $P < 0.001$. Enrichment of two transcription factors each in IGR (left panel) and ORF (right panel) clusters are shown. As expected, Rap1, a known regulator of ribosomal protein genes, is significantly enriched in IGR clusters 1, 2, 3, and ORF cluster 1, which contain genes involved in protein synthesis. The dotted red line is the overall genomic percentage of promoters that are bound by the indicated transcription factor.

response) element which is found upstream of stress-response genes, consistent with the activation of the genes in this cluster in response to various environmental insults (Figure 3D). Presence of *cis*-regulatory DNA elements that are specific to a single cluster may help establish similar patterns of acetylation and contribute to coregulation of the genes within an acetylation cluster.

Acetylation Clusters Are Enriched for Specific Transcription Factor Binding

We also asked whether acetylation clusters are targeted by specific transcription factors using the published binding data for 25 such factors (Lee et al., 2002). Of the 53 IGR clusters, 14 were significantly enriched for 10 different transcription factors at $P < 0.001$. For ORFs, 4 of 68 clusters were enriched on their promoter regions for four different transcription factors at $P < 0.001$. As shown in Figure 5, for example, a significant fraction of genes in IGR clusters 1, 2, 3, and ORF cluster 1 is enriched for the Rap1 transcription factor. These clusters indeed contain genes (e.g., ribosomal protein genes) that are known targets of Rap1 and contain the Rap1 binding element (see above). IGR clusters 45 and 49 are enriched for Cin5, a basic-leucine zipper (bZIP) transcription factor. Reb1, a factor which may regulate both RNA polymerase I and II promoters (Wang et al., 1990), is found upstream of genes in ORF cluster 34. Together with the presence of common DNA motifs, enrichment of clusters for specific transcription factors supports the primary hypothesis for their biological coherence.

Acetylation Arrays Determine the Preference of Bdf1 for Acetylation Sites In Vivo

Since both deacetylated and acetylated lysines may regulate binding of factors that interact with chromatin, we asked whether acetylation arrays can be utilized to determine preference of binding proteins to a particular pattern of (de)acetylated lysines. The double bromodomain-containing protein Bdf1 has been shown to preferentially bind hyperacetylated histones H3 and H4 in vitro. To determine the preference of the Bdf1 protein for sites of acetylation in vivo, we first mapped its genome-wide binding sites on IGR and ORF microarrays (see Experimental Procedures) and then compared the binding map to the acetylation data.

Comparing the IGR or ORF Bdf1-HA binding to the individual sites of acetylation, we find a significant positive correlation with all sites of acetylation examined including H2A and H2B sites (Figure 6A), indicating that there is a strong correlation between overall acetylation level of a chromosomal region and Bdf1 binding.

To determine the relative contribution of each lysine acetylation to the correlation, we compared Bdf1-HA binding to the variance-normalized acetylation data. As shown in Figure 6B, the acetylation of H2AK7 shows the most significant positive correlation with Bdf1 binding on IGRs but not ORFs, whereas H3K14/18/23 acetylation show significant positive correlation with Bdf1 binding on both IGRs and ORFs. Strikingly, H4K16, H3K9/27, and H2BK11 acetylation levels are negatively correlated with Bdf1 binding on IGRs. A similar negative correlation between these same sites, with the exception of H3K27, and Bdf1 binding on ORFs is also observed.

Taken together, these data suggest that while Bdf1 binds to regions of the genome that are generally acetylated on most lysines, the relative acetylation levels among those sites may be important for optimal binding. To verify this possibility, we compared Bdf1 binding at the promoters of four randomly selected target genes (identified from the array data) in strains which contain either one wild-type copy of H4 or a mutated copy of the H4 gene. The mutations include K12R (a lysine to arginine mutation mimics the unacetylated state), K16R, or K16Q (a lysine to glutamine mutation mimics the acetylated state) (Figure 6C). Consistent with the in vitro

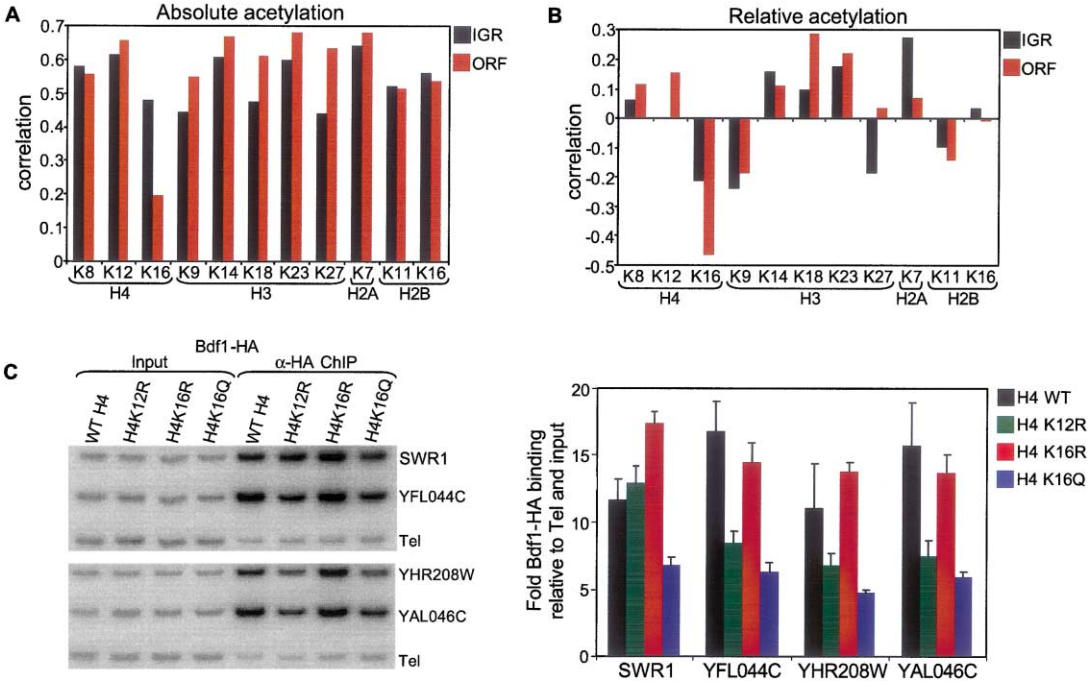

Figure 6. Genome-Wide Bdf1 Binding Correlates with Hypoacetylated H4 Lysine 16

(A) Bdf1 binding shows significant correlations with individual sites of acetylation on IGR (black bars) and ORF (brown bars) regions, indicating that it binds to generally hyperacetylated regions of the genome.

(B) Acetylation levels are first normalized to each other (variance-normalization) and then compared to genome-wide Bdf1 binding, which shows significant positive and negative correlations with acetylation sites.

(C) Assessment of the effect of mutations in histone H4 on Bdf1 binding at the promoter regions of 4 randomly selected Bdf1-target genes in vivo by ChIP. The intensity of Bdf1 enrichment is normalized to a region ~500 bp from the end of chromosome VI-R (Tel) as an internal loading control. The fold binding is the ratio of Tel-normalized values of immunoprecipitated DNA to its input averaged across 3 independent ChIP experiments. A representative gel is shown. K→R, lysine to arginine substitution; K→Q, lysine to glutamine mutation.

data, substituting arginine for lysine at position 12 (K12R) reduces Bdf1 binding by 2–3-fold at the promoters of *YAL046C, YHR208W, and YFL044C* but not *SWR1*. Interestingly, mutating K16 to arginine (K16R) had little effect on Bdf1 binding at the same promoters. However, mutating H4K16 to glutamine (K16Q) reduces Bdf1 binding by 2–4-fold at all promoters examined. The finding that K16Q, but not K16R, mutation reduces binding is consistent with the negative correlation between H4K16 acetylation and Bdf1 binding genome-wide. Thus, when compared to other acetylation sites, hypoacetylation of H4K16 is likely to be required for Bdf1 binding in vivo.

Discussion

Understanding the association of various histone acetylation sites with gene activity requires a global view of the acetylation patterns of individual lysines. Here, we have determined the genome-wide acetylation state of 11 lysines in the four core histones of *Saccharomyces cerevisiae*. We find that combinations of acetylated and deacetylated states of the 11 lysines generate unique patterns of acetylation on promoter as well as coding regions. Clustering of genes with similar patterns of acetylation distinguishes groups of coexpressed genes that are functionally related. In certain cases, these patterns may contribute to the coregulation of genes in the clusters by providing unique binding surfaces on chromatin for transcription regulatory proteins.

Both Hyper- and Hypoacetylation of Histones Are Correlated with Gene Activity

Previous studies have argued that increased histone acetylation is correlated with gene activity. However, this view has been modified by the demonstration that the Hos2 HDAC binds to and deacetylates active genes and is required for gene activity (Wang et al., 2002). A recent study has also implicated Rpd3, normally a repressor, in activation of Hog1 transcription factor target genes (De Nadal et al., 2004). Our global comparisons of the acetylation levels of individual lysines with gene activity now indicates that both acetylation and deacetylation of specific lysines are correlated with transcription. For example, we find that on both intergenic and coding regions genome-wide, hyperacetylation of histone H3K9/18/27 but hypoacetylation of H4K16 and H2BK11/16 are significantly correlated with transcription. In addition, both globally and in the identified clusters, H3K18 and H4K16 are acetylated strikingly in an inverse manner with respect to each other. Thus, the acetylation state of these sites can be a mark for transcriptional status. Indeed this is the case for the ribosomal protein genes, which are coordinately acetylated on H3K18 and deacetylated on H4K16 during logarith-

mic growth when they are highly transcribed. However, these genes become coordinately deacetylated on H3K18 during heat shock when they are repressed (data not shown). Although a causal relationship between H3K18 and H4K16 acetylation remains to be determined, an inverse regulation between acetylated lysines involved in a biological process is intriguing.

Acetylation Clusters Define Groups of Biologically Related Genes

Despite extensive studies of histone (de)acetylation in recent years, it has been unclear whether there exist groups of functionally related genes that are similarly acetylated on either their upstream regions or coding sequences. To address this question systematically and without bias, we applied the *k*-means partitional clustering method to the variance-normalized acetylation data and identified 53 and 68 groups of genes from IGRs and ORFs, respectively. Each IGR or ORF is included in only one cluster, based solely on the similarity of acetylation patterns to other cluster members. Examination of the average acetylation patterns of the IGR clusters reveals a diverse array of acetylation patterns on the promoter (IGR) regions. Remarkably, independent clustering of the ORF acetylation data also uncovered groups of genes with patterns as diverse as those of IGRs.

We show by four different criteria that the genes in the acetylation clusters are biologically related to each other. (1) There is significant coexpression of genes in the clusters. (2) There is significant grouping of genes within the same functional classes or physiological processes. (3) There are upstream DNA sequence motifs that are common and specific to members of each cluster. (4) There is significant enrichment for binding of transcription factors. Not all clusters were significant as assessed by these criteria. The members of such clusters may function in multiple but related processes and so may not show enrichment in any one functional category as defined by current databases or be enriched for a *cis*-regulatory element/transcription factor binding. Interestingly, of the 2206 IGR acetylation data points, 49 are assigned to ORFs that have been designated as dubious ORFs, of which only six fall into five different clusters. Of the 2403 ORFs, 30 are designated as dubious and none falls into the ORF clusters. The exclusion of dubious ORFs by the clustering method provides further support for significant grouping of acetylation clusters for gene classes with different biological characteristics such as function or localization.

Note that some of the clusters have similar levels of expression but distinct patterns of acetylation. For instance, the IGR clusters 29 and 36 have average expression values of 0.26 and 0.27 in log phase, respectively (Figure 3A). Cluster 29 is predominantly deacetylated on H3K9 and acetylated on H2AK7. In contrast, cluster 36 is mainly deacetylated on H3K14, K18, and K23 and acetylated on H4K8, K16, and H2AK7 relative to other lysines (Figure S2). So, both clusters are acetylated on H2AK7 but differ on other sites. While the two clusters are repressed to the same extent in log phase, in other conditions such as stationary phase, the genes in cluster 29 remain repressed but those in cluster 36 are upregulated (Figure S2). Thus acetylation clusters can distinguish groups of genes that have similar expression levels in one condition but are differentially expressed in other conditions. Considering that in both clusters H2AK7 is hyperacetylated, we suggest that patterns of acetylation, rather than individual lysines, may be better correlated with transcriptional status. It is likely that the patterns of acetylation that we have identified in the log phase would be different in other circumstances, such as during cell cycle or under stress conditions. Determining to what extent these patterns may change in other conditions would allow for better understanding of the roles of acetylation sites, individually or in combination, in the regulation of gene activity.

How Are the Acetylation Patterns Established?

Establishment of the acetylation patterns on promoters (IGRs) may involve recruitment of specific sets of HATs and HDACs with different specificities by sequence-specific DNA binding proteins. Sequential or concurrent action of a specific set of these enzymes on the promoter-associated histones could generate unique patterns of acetylation that would be common to any other promoter to which they are recruited. This is supported, for example, by the enrichment of acetylation clusters that consist of ribosomal protein genes for the Rap1 transcription factor. Rap1 is required for the coordinate recruitment of Esa1 HAT (Reid et al., 2000) and Rpd3 HDAC (S.K.K and M.G., unpublished data) to the promoters of ribosomal protein genes.

The mechanism(s) by which ORF acetylation patterns may be established are less clear. The ORF patterns may be determined by the promoter bound regulatory factors. Indeed, we find *cis*-regulatory elements that are common to the upstream regions of the genes within the ORF clusters. The DNA motifs in the promoter regions may serve as binding sites for transcriptional regulators that could direct recruitment of HATs and HDACs to the downstream ORFs, potentially involving the transcription apparatus itself. Transcriptional elongation by RNA polymerase II (RNAPII) requires certain accessory proteins that associate with the polymerase at or near the promoter but remain with the elongating enzyme on the coding region and may facilitate the passage of RNAPII through coding-region nucleosomes. These accessory proteins include the Elp3 HAT as part of the Elongator complex, which when deleted (along with Gcn5 HAT) results in preferential hypo-acetylation of coding regions (Kristjuhan et al., 2002). The Hos2 HDAC also associates with and deacetylates the coding region nucleosomes only during gene activity (Wang et al., 2002). Therefore, association of HATs and HDACs with the elongating form of RNAPII provides a mechanism for acetylation and deacetylation of nucleosomes restricted to the coding regions.

Acetylation Status of Histone H4 Lysine 16 Regulates Protein Binding to Chromatin

Bdf1 protein contains bromodomains that recognize acetylated histone H3 and H4 tails in vitro. Our data indicate that Bdf1 may have an expanded specificity in vivo. In support of this, we find that mutation of not only H4K12R disrupts Bdf1 binding in vivo (consistent with the in vitro data), but also H2AK7R as well as dele-

tion of residues 4–20 of the H2A N terminus, which significantly reduces Bdf1 binding genome-wide (data not shown). This analysis, however, cannot rule out an indirect effect of the histone mutations on binding of other proteins that in turn affect Bdf1 binding. Significantly, Bdf1 binding correlates with relative hypoacetylation of H4K16 both on IGRs and ORFs and requires the presence of a positive charge at position 16 of H4 at several promoters. Altogether, the data suggest that Bdf1 binds to generally hyperacetylated chromatin in vivo and that hypoacetylation of H4K16 may facilitate this interaction.

Interestingly, the H4 N-terminal tail that is hypoacetylated uniquely at K16 is also a binding target for heterochromatin protein Sir3 in yeast and the chromatin-remodeling protein ISWI in flies. Acetylation of H4K16 disrupts H4-Sir3 interaction both in vitro and in vivo (Kimura et al., 2002; Suka et al., 2002) and inhibits substrate recognition by ISWI (Corona et al., 2002). In fact, H4K16 hypoacetylation may not only be required for association of proteins with heterochromatin (Sir3) and repressed chromatin (ISWI), but also with euchromatin (Bdf1). Thus, the acetylation status of H4K16 may be a general mechanism for regulation of binding of both positive and negative regulators of transcription.

Do the Acetylation Patterns Constitute a Code?

The diversity of histone modifications and their recognition by other proteins have led to the proposal that distinct histone marks "…act sequentially or in combination to form a 'histone code' that is read by other proteins to bring about distinct downstream events" (Strahl and Allis, 2000). We believe that the use of histone modifications individually or sequentially cannot be considered a code since the total number of modifications do not necessarily contain more information than the sum of individual modifications (Kurdistani and Grunstein, 2003; Turner, 2002). Single or sequential use of the modifications is characteristically similar to a protein-signaling pathway (Schreiber and Bernstein, 2002). A "histone code" must be both combinatorial and consistent for a particular output (e.g., protein binding, transcriptional activity). To date, the use of acetylation of individual lysines has been shown to be neither combinatorial nor consistent from gene to gene. Our data now provide some evidence for groups of genes with consistent acetylation patterns involving combinations of 11 lysines. Still such patterns may represent in part the culmination of cascading mechanisms for successive acetylation and deacetylation of individual lysines, with each event specifying a unique output. This would not constitute a code. Only if the acetylation patterns are read combinatorially, can they provide the language for a true code.

In any case, it will be of interest to determine to what extent different patterns of histone modifications are important for regulation of not only transcription but other chromosomal processes such as DNA replication, repair, recombination, etc. in which histone acetylation has been shown to play a role.

Experimental Procedures

Yeast Strains

All data were obtained from the YDS2 wild-type (wt) yeast strain. We tagged the *BDF1* gene with three copies of HA in RMY200 (wt H4), NSY268 (H4K12R), JTY501U (H4K16R), and JTY502U (H4K16Q) by PCR-based tagging of the 3′ end in the chromosomal locus through homologous recombination using the plasmids as described (Longtine et al., 1998) to generate SKY150, SKY151, SKY152, and SKY153, respectively.

Chromatin Immunoprecipitation and Microarrays

Chromatin immunoprecipitation (ChIP) was performed essentially as described (Suka et al., 2001) on yeast cells grown to A600 of ~1 in 50 ml YEPD medium, using 2–5 μl (per 50 μl lysate) of highly specific antibodies raised against individual sites of acetylation. For Bdf1-HA ChIP, the 12CA5 monoclonal anti-HA antibody (0.5 μl per 50 μl lysate) was used. DNA microarray experiments were performed as previously described (Robyr et al., 2002). Arrays containing intergenic regions were produced in house as reported, and those containing open reading frames (ORFs) were purchased from University Health Networks, Toronto. Microarrays were scanned (Agilent Array Scanner) and fluorescent intensities were quantified using the Imagene software (Biodiscovery).

Normalization and Data Analysis

The genome-wide measurements of acetylation for the 11 residues were all normalized to have an average value of 1 across the genome. Thus, hypoacetylated regions may in fact be acetylated but to a lower degree. This normalization was performed to address differences in values obtained for each lysine due to differential efficiency of each antibody in ChIP. The acetylation values for H4K8 were also normalized using the Lowess method (Yang et al., 2002), which yielded essentially identical results as those obtained from total-intensity normalization for both IGRs ($r = 0.96$) and ORFs ($r = 0.99$). To determine the relative extent of acetylation at each residue, we performed a variance-normalization procedure (typically used in clustering) across the 11 residues for each locus. This is done by subtracting the mean acetylation level for all the measured lysines from individual measurements and dividing by their standard deviation. In this way, each locus has a mean acetylation level of zero and variance of 1. Each of these variance-normalized values then corresponds to number of standard deviations above (positive) or below (negative) the mean. Other relative measures gave similar results.

Clustering

We used a slightly modified *k*-means algorithm to find clusters of genes with similar acetylation patterns across the 11 residues. Each cluster was constrained to only include genes with some minimal cut-off Pearson correlation coefficient (in our case, 0.70), and to have more than a minimal number of genes (in our case, 20). Any cluster which does not satisfy this size constraint is reseeded from a random gene (Beer and Tavazoie, 2004).

Significant mRNA Expression Coherence of Acetylation Clusters

To determine whether the members of an acetylation cluster were significantly correlated in their pattern of mRNA expression, we calculated all pair-wise Pearson correlation coefficients for the expression patterns of all the genes within a cluster. For a cluster with *n* members, this corresponds to *n(n-1)/2* correlation values. We define "expression coherence" as the median of these values. To determine whether the members of an acetylation cluster have significant expression coherence, we performed N simulations, each of which consisted of picking *n* random genes and measuring their expression coherence. If in none of these N random simulations, the expression coherence exceeded the actual value for the cluster, we deemed the cluster's expression coherence significant with $P < 1/N$.

Significant Extremes in Average Expression Level of Acetylation Clusters

To determine whether the members of each acetylation cluster had either unusually high or low average expression level, we used a published mRNA expression dataset gathered under similar conditions as the acetylation measurements were performed (Causton et al., 2001). We used the average intensity values for the members of

each cluster as a measure of average expression level for each cluster. To determine significantly high/low average expression level, we performed random simulations as above, but considered both high and low extremes to determine significance.

Enrichment of Acetylation Clusters for Biologically Coherent Characteristics

To determine whether the members of each acetylation cluster were significantly enriched in any of known categories of biological function, fitness, cellular compartment, etc., we used the FunSpec website (Functional Specifications at http://funspec.med.utoronto.ca) which contains a comprehensive database of these features, and performs statistical tests using the hypergeometric distribution to evaluate significant enrichment. All reported values in Supplemental Table S1 (available on *Cell* website) are corrected for multiple testing using the conservative Bonferroni correction.

Motif Finding

We searched for overrepresented DNA sequence motifs upstream of genes within each acetylation cluster as previously described (Hughes et al., 2000; Tavazoie et al., 1999). We used the AlignACE algorithm with standard *S. cerevisiae* parameters to look for significant motifs up to 800 bp upstream of each gene. Motif finding was performed for the genes in IGR and ORF acetylation clusters. The roughly $\sim$2000 motifs generated contained much redundancy and were therefore clustered by motif similarity using the CompareACE algorithm (Hughes et al., 2000). Representative members of each motif cluster were combined with a set of 49 known TF binding site motifs to form a final set of 354 motifs, which were used in subsequent analysis.

Motif Occurrence and ChIP-Derived Binding of Transcription Factors

Significant enrichment of acetylation clusters for each motif/TF binding was assessed by performing N simulations with randomly permuted matrices. If the prevalence of motif/TF binding in these random simulations did not exceed the actually observed values, the motif/TF binding was deemed to be significantly associated with a cluster with $p < 1/N$. The P values reported as significant were not corrected for multiple testing, but with a conservative Bonferroni correction, they were all significant with $P < 0.05$.

Acknowledgments

We are grateful to members of the Grunstein and Tavazoie laboratories for critical comments and discussion throughout this work. S.K.K. is a Howard Hughes Medical Institute Physician Postdoctoral Fellow. This work was supported by Public Service grants of NIH to M.G. and by National Science Foundation CAREER Award and grants from DARPA and DOE to S.T.

Received: February 26, 2004
Revised: May 25, 2004
Accepted: May 26, 2004
Published: June 10, 2004

References

Agalioti, T., Chen, G., and Thanos, D. (2002). Deciphering the transcriptional histone acetylation code for a human gene. Cell *111*, 381–392.

Beer, M.A. and Tavazoie, S. (2004). Predicting gene expression from sequence. Cell *117*, 185–198.

Causton, H.C., Ren, B., Koh, S.S., Harbison, C.T., Kanin, E., Jennings, E.G., Lee, T.I., True, H.L., Lander, E.S., and Young, R.A. (2001). Remodeling of yeast genome expression in response to environmental changes. Mol. Biol. Cell *12*, 323–337.

Corona, D.F., Clapier, C.R., Becker, P.B., and Tamkun, J.W. (2002). Modulation of ISWI function by site-specific histone acetylation. EMBO Rep. *3*, 242–247.

De Nadal, E., Zapater, M., Alepuz, P.M., Sumoy, L., Mas, G., and Posas, F. (2004). The MAPK Hog1 recruits Rpd3 histone deacetylase to activate osmoresponsive genes. Nature *427*, 370–374.

Deuring, R., Fanti, L., Armstrong, J.A., Sarte, M., Papoulas, O., Prestel, M., Daubresse, G., Verardo, M., Moseley, S.L., Berloco, M., et al. (2000). The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. Mol. Cell *5*, 355–365.

Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., and Zhou, M.M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. Nature *399*, 491–496.

Hartigan, J. (1975). Clustering algorithms (New York, NY: Wiley).

Horn, P.J., and Peterson, C.L. (2002). Molecular biology. Chromatin higher order folding–wrapping up transcription. Science *297*, 1824–1827.

Hughes, J.D., Estep, P.W., Tavazoie, S., and Church, G.M. (2000). Computational identification of cis-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*. J. Mol. Biol. *296*, 1205–1214.

Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science *288*, 1422–1425.

Kimura, A., Umehara, T., and Horikoshi, M. (2002). Chromosomal gradient of histone acetylation established by Sas2p and Sir2p functions as a shield against gene silencing. Nat. Genet. *32*, 370–377.

Kristjuhan, A., Walker, J., Suka, N., Grunstein, M., Roberts, D., Cairns, B.R., and Svejstrup, J.Q. (2002). Transcriptional inhibition of genes with severe histone h3 hypoacetylation in the coding region. Mol. Cell *10*, 925–933.

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

Ladurner, A.G., Inouye, C., Jain, R., and Tjian, R. (2003). Bromodomains mediate an acetyl-histone encoded antisilencing function at heterochromatin boundaries. Mol. Cell *11*, 365–376.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., et al. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science *298*, 799–804.

Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953–961.

Ma, X.J., Wu, J., Altheim, B.A., Schultz, M.C., and Grunstein, M. (1998). Deposition-related sites K5/K12 in histone H4 are not required for nucleosome deposition in yeast. Proc. Natl. Acad. Sci. USA *95*, 6693–6698.

Mann, R.K., and Grunstein, M. (1992). Histone H3 N-terminal mutations allow hyperactivation of the yeast GAL1 gene in vivo. EMBO J. *11*, 3297–3306.

Matangkasombut, O., and Buratowski, S. (2003). Different sensitivities of bromodomain factors 1 and 2 to histone H4 acetylation. Mol. Cell *11*, 353–363.

Reid, J.L., Iyer, V.R., Brown, P.O., and Struhl, K. (2000). Coordinate regulation of yeast ribosomal protein genes is associated with targeted recruitment of Esa1 histone acetylase. Mol. Cell *6*, 1297–1307.

Robyr, D., Suka, Y., Xenarios, I., Kurdistani, S.K., Wang, A., Suka, N., and Grunstein, M. (2002). Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. Cell *109*, 437–446.

Schreiber, S.L., and Bernstein, B.E. (2002). Signaling network model of chromatin. Cell *111*, 771–778.

Sobel, R.E., Cook, R.G., Perry, C.A., Annunziato, A.T., and Allis, C.D. (1995). Conservation of deposition-related acetylation sites in newly synthesized histones H3 and H4. Proc. Natl. Acad. Sci. USA *92*, 1237–1241.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41–45.

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

Suka, N., Luo, K., and Grunstein, M. (2002). Sir2p and Sas2p oppositely regulate acetylation of yeast histone H4 lysine16 and spreading of heterochromatin. Nat. Genet. *32*, 378–383.

Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., and Church, G.M. (1999). Systematic determination of genetic network architecture. Nat. Genet. *22*, 281–285.

Tse, C., Sera, T., Wolffe, A.P., and Hansen, J.C. (1998). Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. Mol. Cell. Biol. *18*, 4629–4638.

Turner, B.M. (2002). Cellular memory and the histone code. Cell *111*, 285–291.

Wang, H., Nicholson, P.R., and Stillman, D.J. (1990). Identification of a Saccharomyces cerevisiae DNA-binding protein involved in transcriptional regulation. Mol. Cell. Biol. *10*, 1743–1753.

Wang, A., Kurdistani, S., and Grunstein, M. (2002). Requirement of Hos2 histone deacetylase for gene activity in yeast. Science *298*, 1412–1414.

Wu, J., Suka, N., Carlson, M., and Grunstein, M. (2001). TUP1 utilizes histone H3/H2B-specific HDA1 deacetylase to repress gene activity in yeast. Mol. Cell *7*, 117–126.

Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., and Speed, T.P. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res. *30*, e15.