ANALYSIS

nature genetics

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

A genome-wide analysis in *Saccharomyces cerevisiae* demonstrates the influence of chromatin modifiers on transcription

Israel Steinfeld[1,2], Ron Shamir[2] & Martin Kupiec[1]

Chromatin structure is important in transcription regulation. Many factors influencing chromatin structure have been identified, but the transcriptional programs in which they participate are still poorly understood. Chromatin modifiers participate in transcriptional control together with DNA-bound transcription factors. High-throughput experimental methods allow the genome-wide identification of binding sites for transcription factors as well as quantification of gene expression under various environmental and genetic conditions. We have developed a new methodology that uses the vast amount of available data to dissect the contribution of chromatin structure to transcription. We measure and characterize the dependence of transcription factor function on specific chromatin modifiers. We apply our methodology to *S. cerevisiae*, using a compendium of 170 gene expression profiles of strains defective for chromatin modifiers, taken from 26 different studies. Our method succeeds in identifying known intricate genetic interactions between chromatin modifiers and transcription factors and uncovers many previously unknown genetic interactions, giving the first genome-wide picture of the contribution of chromatin structure to transcription in a eukaryote.

The fate of a given cell is determined by its particular program of gene expression. In eukaryotic genomes, gene regulation at the transcriptional level is governed mainly by proteins that facilitate transcription by binding to gene promoters and either recruiting or preventing the recruitment of the transcription machinery. In this paper, we will refer to both types of regulators (positive and negative) by the general term 'transcription factor'. New experimental techniques allow *in vivo* genome-wide mapping of transcription factor binding[1]. The recent development of technologies that can characterize the group of genes targeted (bound) by the transcription factor facilitates research on the regulatory forces imposed on the transcription factor.

The efficiency of a transcription factor in governing transcription depends on its affinity for the promoters of the group of genes it binds (hereafter called the transcription factor 'cohort'). Chromatin configuration may also determine the accessibility of the promoter to external factors and the performance of the transcription machinery[2,3]. Chromatin modifiers influence chromatin structure by enabling the formation of a chromatin structure needed for transcription factor activity (**Fig. 1a**). Chromatin modifiers can be factors that use ATP, or they can act independently of ATP. Among the ATP-independent chromatin modifiers, a widely explored group comprises the histone acetyltransferases (HATs) and the histone deacetylases (HDACs)[2]. The addition of acetyl groups to specific lysine residues on the N-terminal histone tails by the HATs is believed to create a less condensed chromatin structure. Previous work has shown that hyperacetylated regions are in general highly transcribed, whereas hypoacetylated regions are silent[4]. Other ATP-independent chromatin modifiers modify histones by adding methyl groups, phosphates or ubiquitin moieties. The mechanisms by which these modifications affect transcription constitute one of the most active areas of current research. The ATP-dependent chromatin remodelers usually act as multiprotein complexes that contain an ATPase subunit[3], although the mechanism by which they act is still unclear. Some remodelers are able to destabilize the nucleosomes, allowing the binding of factors to the DNA; others can shift the position of nucleosomes along the chromatin, affecting chromatin structure[2].

Some transcription factors require the recruitment of a chromatin modifier to facilitate their activity (**Fig. 1b**). In such cases, the chromatin modifier could be seen as a cofactor of transcription. The budding yeast *S. cerevisiae* is an excellent organism for modeling eukaryotic transcription regulation, and interactions between particular chromatin modifiers and transcription factors have been studied in detail in this organism[5,6].

We have assembled a large compendium of gene expression experiments in which various chromatin modifiers were deleted or genetically manipulated. Using a statistical approach, we have carried out a systematic search for transcription factor–chromatin modifier pairs that function in concert. Our compendium allows a system-level overview of the effect of chromatin on transcription and also pinpoints specific transcription factor–chromatin modifier interactions.

The chromatin modifier compendium

We used two types of data in this work. For each transcription factor, we selected the group of genes it binds (its cohort) based on a genome-wide transcription factor–DNA binding experiment[1]. By applying a strict binding threshold ($P < 0.001$), we ensured a low level of false

[1]Department of Molecular Microbiology and Biotechnology and [2]School of Computer Science, Tel Aviv University, Ramat Aviv 69978, Israel. Correspondence should be addressed to M.K. (martin@post.tau.ac.il).
Published online 26 February 2007; doi:10.1038/ng1965

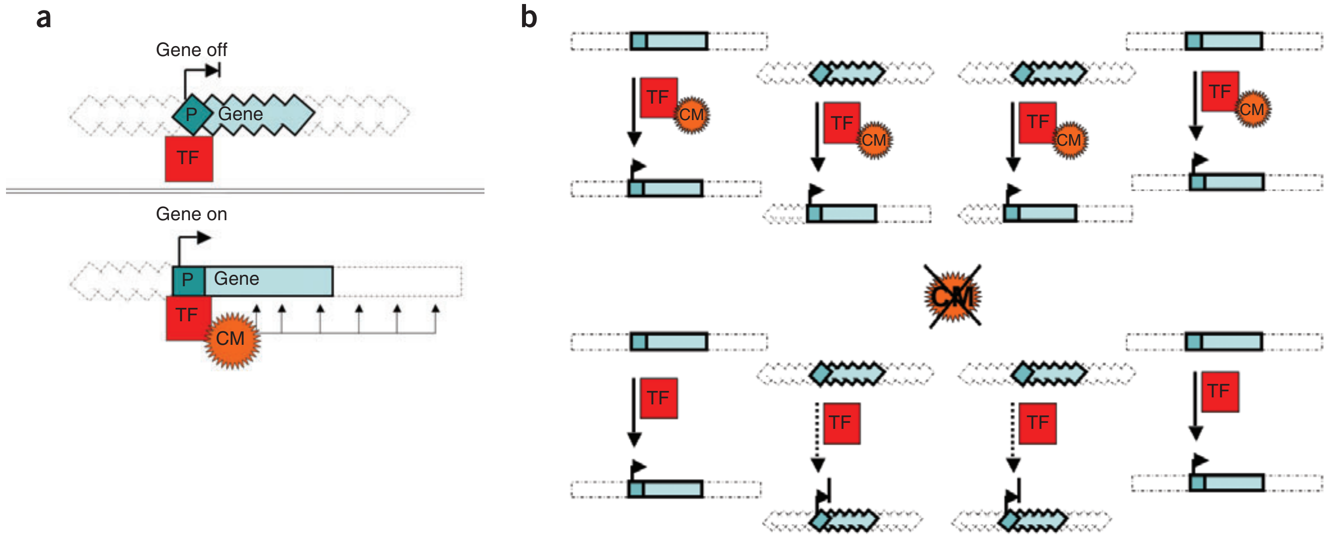

Figure 1 A model for chromatin modifier-mediated transcription. (**a**) In a 'closed' chromatin structure (top), the transcriptional machinery is less accessible to the gene's promoter, diminishing the efficiency of transcription. The activity of a chromatin modifier relaxes the chromatin into an 'open' structure (bottom), which promotes transcription by facilitating accessibility. In some cases, chromatin modifiers are known to act in the opposite direction, causing the chromatin to adopt a more compact configuration and thus preventing gene expression[3,4,24]. (**b**) Upon activation of a transcription factor, each of the transcription factor target genes is induced according to various parameters, including its chromatin structure, among others. The interaction between transcription factor and chromatin modifier enables the activation of genes located in regions with 'closed' chromatin (top). In strains with mutant chromatin modifier genes, the absence of the chromatin modifier will lead to changes in the expression of genes that depend on the chromatin modifier for transcription (bottom).

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

positives (<8%)[7]. We gathered the second data set, a gene expression compendium, from the literature; it contains experiments carried out with yeast strains in which genes encoding particular chromatin modifiers were mutated (**Supplementary Table 1** online). This compendium, consisting of 170 gene expression profiles taken from 26 different publications, covers more than 60 potential interacting chromatin modifiers such as HATs (the NuA4, HAT1 and SAGA complexes), HDACs (the RPD3, HDA1 and SET3 complexes), histone methyltransferases (the COMPASS complex), ATP-dependent chromatin remodelers (the SWI/SNF, SWR1, INO80, ISWI and RSC complexes) and other chromatin-affecting genes and cofactors such as Spt10, Sir proteins and the TATA-binding protein (TBP) (**Fig. 2**). Hence, this collection provides a valuable tool for analyzing the involvement of chromatin modifiers in transcription.

The rationale of our work was as follows: mutations in a gene encoding a given chromatin modifier affect transcription of many genes. If, however, regulation by a specific transcription factor depends on the activity of a particular chromatin modifier, we expect that those mutations will cause a preferential effect on expression of the transcription factor target genes (**Fig. 1b**). For each strain with a mutated chromatin modifier, we partitioned gene expression profiles into two groups: the transcription factor cohort and the rest of the genes. If the transcription factor and chromatin modifier cooperate in controlling the expression of a subset of genes, deletion of the chromatin modifier should cause a differential change in expression (**Fig. 1b**). To evaluate the difference in the distribution of gene expression values in the two groups (the cohort and the control), we used the Kolmogorov-Smirnov (K-S) statistical test[8]. The K-S *P* value provides a measure of the discrepancy in expression between the transcription factor cohort and the rest of the genes when the chromatin modifier gene is mutated. The K-S score expresses both the direction and significance of the disparity between the two distributions. Positive scores indicate activation, whereas negative scores imply reduced expression of the transcription factor cohort (see Methods).

Ume6 regulation

We first tested our method on the well-characterized example of the transcription factor Ume6, a central regulator of early meiotic genes that is known to regulate its cohort through interactions with chromatin modifiers[5]. During vegetative growth, binding of Ume6 upstream of specific early meiotic genes facilitates the recruitment of the RPD3 complex (an HDAC) and Isw2 (an ATP-dependent chromatin remodeler)[9,10]. Rpd3 has been shown to deacetylate histones H3 and H4 (ref. 11), and this hypoacetylation, along with the recruitment of Isw2, is presumed to create a condensed chromatin structure that prevents gene expression[9]. During entry to meiosis, Ume6 preferentially interacts with the activator Ime1 to promote expression of its meiosis-related cohort[12].

The Ume6 cohort, as defined in ref. 1, consists of 131 genes. As expected, deletion of *UME6* leads to a significant shift in the expression pattern of the Ume6 cohort (**Table 1** and **Fig. 3**). Ume6 acts as a repressor only through its ability to recruit the Rpd3 complex and the Isw2 chromatin remodeler to its binding location[9]. According to this dogma, not only deletion of *UME6* but also double deletion of *ISW2* and *RPD3* should derepress all Ume6-regulated genes. Our results show exactly this effect: the Ume6 cohort showed a significant activation in an experiment carried out with the doubly deleted *isw2*Δ *rpd3*Δ strain

Table 1 Deviation of the Ume6 cohort in various chromatin modifier gene expression experiments

Publication	K-S score	Condition
Ref. 15	**13.02**	ume6Δ
Ref. 13	**11.29**	rpd3Δ H3ΔN vs. H3ΔN
Ref. 46	**7.54**	ume6Δ
Ref. 15	**6.7**	isw2Δ rpd3Δ
Ref. 46	**5.48**	rpd3Δ
Ref. 15	–0.36	isw2Δ

Boldface indicates significant K-S scores.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

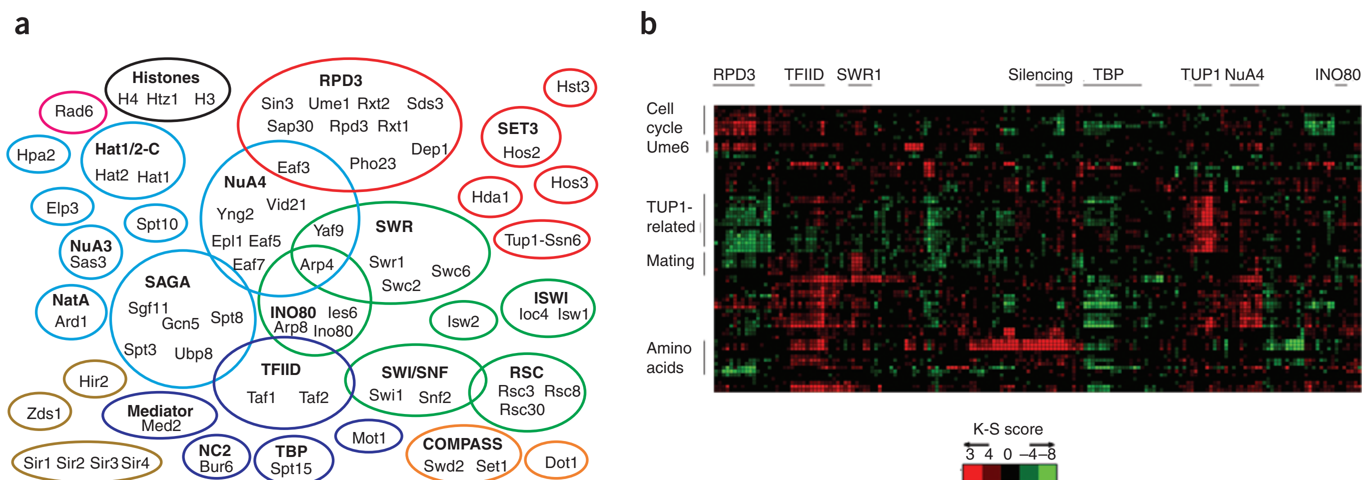

Figure 2 The chromatin modifier gene expression compendium. (**a**) The expression profiles available in the compendium. Each of the listed chromatin modifiers has one or more profiles in the compendium, created by a genetic alteration of the chromatin modifier in the yeast genome. Chromatin modifiers that belong to the same complex are circumscribed by an oval, with the complex name in bold. Colors indicate the chromatin modifier's proposed biochemical activity: HATs in light blue, HDACs in red, methyltransferases in orange, ubiquitin-conjugating enzymes in magenta, chromatin remodelers in green, TAF-related factors in dark blue, silencing factors in brown and histone subunits in black. (**b**) Clustering of the compendium. Rows represent transcription factor cohorts, and columns represent conditions. Colors indicate chromatin modifier cohort K-S scores. To obtain a global view of the transcription factor–chromatin modifier interaction landscape, we hierarchically clustered the cohorts and conditions according to their K-S scores (positive scores in red and negative in green). Groups of functionally related transcription factors (ordinate) and functionally related conditions (abscissa) are marked. The detailed hierarchical clustering solution is available in **Supplementary Figure 1**.

(**Fig. 3b**). As predicted, we saw a less significant effect for strains with individual deletions of *RPD3* or *ISW2* (**Table 1**). The effect is not due to a reduction in expression of the *UME6* gene itself (**Supplementary Table 2** online). Examination of the derepressed genes (*Z* score > 1) from the Ume6 cohort in both the *ume6*Δ and the *isw2*Δ *rpd3*Δ experiments uncovered a significant overlap (hypergeometric $P < 4.6 \times 10^{-7}$) (**Fig. 4a**). The similar effect observed in both experiments points to the common mechanism of regulation by Ume6 and Isw2 with Rpd3.

Reassured by the ability of our methodology to expose the well-characterized contribution of Isw2 and Rpd3 to Ume6 regulation, we carried out a systematic exploration of the Ume6 cohort in the entire compendium. Others[13] have explored the relationship between the transcription regulation by Rpd3 and the N termini of histones H3 and H4. Because the deletion of the N-terminal domain of histones prevents their regulation by most ATP-independent chromatin modifiers, strains were constructed carrying mutant versions of either histone H3 or histone H4 in which the N terminus of the protein was deleted (*H3ΔN* and *H4ΔN*, respectively)[13]. To test whether Rpd3 has an effect on gene expression independent of H3, we compared the *H3ΔN* strain for which *RPD3* was also deleted with the isogenic *H3ΔN* strain. This showed a highly significant and specific disparity in the expression of the Ume6 cohort (**Fig. 3c**). The activated genes from the Ume6 cohort (*Z* score > 1) in this experiment shared a significant overlap with those derepressed in the strain with deletion of *UME6*, as well as with the strain doubly deleted for *ISW2* and *RPD3* (hypergeometric $P < 10^{-3}$ and $P < 10^{-4}$, respectively) (**Fig. 4a**). Notably, in the parallel experiment carried out with *H4ΔN*, we did not observe any effect (**Fig. 3d**). *In vitro* studies have implicated both the H3 and H4 histones in the binding of *ISW2* to nucleosomes[14,15]. The additive effect of the *RPD3* deletion to the mutation in the gene encoding H3, as opposed to the mutation in the gene encoding H4, suggests that histone H4, but not H3, is likely to work with Rpd3. In addition, the similar effects obtained in the *rpd3*Δ strain lacking the N terminus of histone H3 and in the *rpd3*Δ strain lacking *ISW2* suggest that H3 tails are central in the recruitment of Isw2 by Ume6. Hence, our method enabled the discovery of known Ume6 chromatin modifier cofactors solely by exploring the behavior of the Ume6 cohort in various experiments.

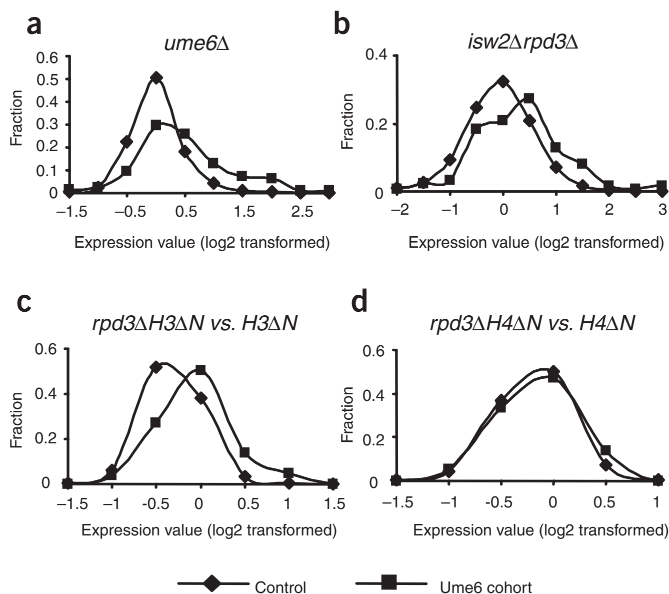

Figure 3 Distribution of expression values for the Ume6 cohort in various chromatin modifier knockout experiments. Distributions of expression levels (log2 transformed) are presented for the Ume6 cohort and the control group (rest of the genes). (**a**) Strain with *UME6* deletion (ref. 9). (**b**) Strain with deletion of *ISW2* and *RPD3* (ref. 9). (**c**) Strain deleted for *RPD3* along with a deleted N terminus of histone H3 compared with an isogenic strain carrying only the histone mutation[13]. (**d**) Strain deleted for *RPD3* along with a deleted N terminus of histone H4 compared with an isogenic strain carrying only the histone mutation[13].

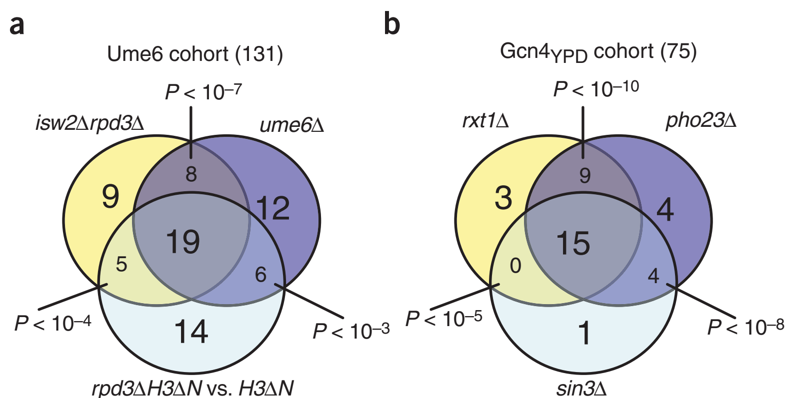

Figure 4 Overlap in altered cohort genes. Level of overlap between altered cohorts in various gene expression experiments (see Methods). (**a**) Overlap in derepressed Ume6 cohort genes in three experiments. Out of 131 Ume6 cohort genes, 45 showed a notable induction (*Z* score > 1) in a *UME6* deleted strain[9], 41 in a doubly deleted *ISW2 RPD3* strain[9] and 44 in strain deleted for *RPD3* along with a deleted N terminus of histone H3 compared with an isogenic strain carrying only the histone mutation[13]. The significance of the overlap between each pair of strains is indicated (hypergeometric *P* value). (**b**) Overlap in activated $Gcn4_{YPD}$ cohort genes in three experiments. Out of 75 $Gcn4_{YPD}$ cohort genes, 32 showed a notable induction in a strain with deletion of *PHO23*, 27 in a strain with deletion of *RXT1* and 20 in a strain with deletion of *SIN3*[21].

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

Systematic exploration of the compendium

The method described above can be applied to any transcription factor with a sufficiently large known cohort. Out of the 204 transcription factors analyzed in ref. 1, 49 generated cohorts large enough to continue with the analysis; of these, 19 were analyzed in more than one environment. In total, we were able to analyze 75 cohorts (see Methods). Our test generated 4,645 transcription factor–chromatin modifier pairs with a K-S $P < 0.05$, and after correction for multiple testing, 531 significant pairs remained (|K-S score| > 5.41; see Methods) (**Supplementary Table 2**). The significant pairs came from 55 different cohorts (defined for 35 transcription factors) and 129 gene expression experiments, covering most of the complexes known to participate in chromatin structure regulation (**Fig. 2**). In total, we obtained 287 unique pairs (one transcription factor and one chromatin modifier (**Supplementary Table 3** online)), giving the first global picture of the transcription factor contribution in chromatin structure regulation in a eukaryote (**Supplementary Note** online).

We obtained a global view of the compendium and its interplay with the transcription factor cohorts by hierarchical clustering of chromatin modifiers and of transcription factors according to similarity of their K-S score profiles across all experimental conditions and cohorts (**Fig. 2b** and **Supplementary Fig. 1** online). This procedure enabled the visualization of common trends of different cohorts in response to all the chromatin modifier perturbations as well as the detection of chromatin modifiers with similar specificity according to their effect on the cohorts. When clustered, transcription factor cohorts appear to be grouped according to various biological processes: cell cycle, amino acid biosynthesis, mating and more. The inclusion of two transcription factors in the same group is sometimes due to a high level of overlap between their cohorts but in many cases reflects common chromatin modifier–mediated mechanisms of regulation. Our results suggest that the genome is organized along functional similarities and that cohorts involved in common biological processes are affected by similar chromatin modifiers. In the case of cell cycle progression, for example, transcription factors affecting different stages are nonetheless grouped together, implying a common interplay with chromatin modifiers.

Our complete results on chromatin modifier–transcription factor interaction are available in **Supplementary Table 2**. Our analysis uncovered many new putative transcription factor–chromatin modifier interactions. In the following section, we focus on several interesting cases where a mutation in a gene encoding a specific chromatin modifier has a significant effect on a transcription factor cohort.

Gcn4 as a repressor of amino acid biosynthetic genes

The Gcn4 transcription factor activates many genes under conditions of amino acid starvation[16]. Initiation of transcription by Gcn4 depends on many coactivators[6], including the SWI/SNF and SAGA complexes, recruited by Gcn4 in response to amino acid starvation[17,18]. In accordance with the positive role of Gcn4, its cohort was strongly repressed in the expression profile of a *gcn4*Δ strain[19] and strongly activated in a strain overexpressing *GCN4*[20]. Harbison *et al.*[1] defined the Gcn4 cohort in an experiment carried out in rich medium ($Gcn4_{YPD}$) and also in cells exposed to sulfometuron methyl (SM), an inhibitor of several amino acid biosynthesis pathways ($Gcn4_{SM}$). The $Gcn4_{SM}$ cohort is larger and consists of 189 genes, but notably, the $Gcn4_{YPD}$ cohort, which consists of only 75 genes, is a subset of the SM cohort[1], indicating that Gcn4 binds to its core cohort under all growth conditions. Keogh *et al.*[21] thoroughly analyzed the RPD3 complex using biochemical and genetic tools. The authors defined two distinct RPD3 complexes, RPD3(L) and RPD3(S), which share a core of three proteins: Rpd3, Sin3 and Ume1. Eaf3 and Rco1 are unique to the RPD3(S) small complex, whereas Pho23, Rxt1 and Rxt2 are specific to the larger RPD3(L) complex. Our results show a clear activation of the Gcn4 cohort when subunits of the large RPD3(L) complex are deleted (**Table 2**). Gcn4 activation was not due to higher levels of Gcn4 itself (**Supplementary Table 2**). Moreover, the activated genes were highly overlapping in each experiment carried out with strains deleted for RPD3(L) subunits (**Fig. 4b**), emphasizing the

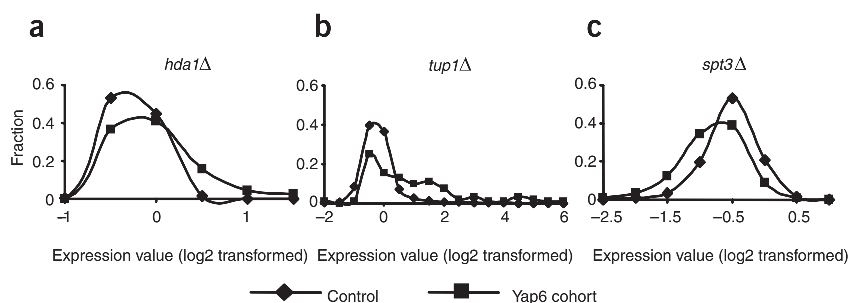

Figure 5 Distribution of expression values for the Yap6 cohort in various chromatin modifier knockout experiments. Distributions of expression levels (log2 transformed) are presented for the Yap6 cohort and the control group (the rest of the genes). Legends are as in **Figure 3**. (**a**) Strain with *HDA1* deleted[45]. (**b**) Strain with *TUP1* deleted[19]. (**c**) Strain with *SPT3* deleted[27].

Table 2 K-S scores for the Gcn4 cohorts in strains deleted for particular RPD3C subunits[21]

$Gcn4_{SM}$ K-S score	$Gcn4_{YPD}$ K-S score	Constituent of	Condition
17.34	**7.73**	RPD3(L)	pho23Δ
13.01	**5.92**	RPD3(L)	rxt1Δ
6.69	4.71	RPD3(L)	rxt2Δ
10.15	4.05	Core complex	sin3Δ
9.16	2.8	Core complex	rpd3Δ
6.42	2.71	Core complex	ume1Δ
0.004	0.06	RPD3(S)	eaf3Δ
0.040	0.022	RPD3(S)	rco1Δ

Boldface indicates significant K-S scores.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

essential contribution of the RPD3(L) complex to the regulation by Gcn4. Notably, when subunits of the RPD3(S) complex (which has been linked to Set2 methyl-transferase[21]) were deleted, the cohort did not show any disparity from the rest of the genes (**Table 2**). Thus, the two complexes also have functionally divergent roles in the regulation by Gcn4.

The gene expression experiments carried out by Keogh *et al.*[21] were not done in amino acid–limiting conditions but rather in rich medium. However, we observed the Rpd3 effect on all the Gcn4 targets. The additional targets available in the $Gcn4_{SM}$ cohort preserved the described trend and even showed stronger activation in the experiments carried out with strains lacking some RPD3(L) subunits but not in strains lacking some RPD3(S) subunits. This effect points to a wide participation of the Rpd3 complex in the regulation by Gcn4, an effect seen even on weak targets of Gcn4 in rich medium. Gcn4 has been shown to use SAGA, a histone acetyltransferase, to activate its cohort[6]. Our results point to the opposite biochemical reaction, removal of acetyl groups from histones, performed by the RPD3 HDAC complex, as a mechanism that can maintain its target genes in an inactive state. Functional analysis on the activated genes in the experiments in which RPD3(L) members were deleted showed an overrepresentation of arginine biosynthesis genes (all eight genes involved in arginine biosynthesis present in the in $Gcn4_{YPD}$ cohort showed increased expression; $P < 0.001$). Thus, our results suggest that Rpd3 and Gcn4 act as negative regulators of the arginine biosynthesis pathway under optimal growth conditions.

Regulation of Yap6 through repression by Tup1

Having tested our methodology on the well-characterized example of Gcn4, we asked whether we could find novel interactions for less-characterized transcription factors. Very little is known about the transcription factor Yap6; it has sequence similarity to AP-1 and has been linked to lithium and sodium resistance[22]. Examination of the behavior of the Yap6 cohort against the entire compendium demonstrated a range of potential interactions with various chromatin modifiers (**Supplementary Table 2**), which is surprising, given the anonymity of Yap6. We observed a significant activation of the Yap6 cohort in an HDA1-null strain (**Fig. 5a**). Hda1 is the catalytic member of the *HDA1* HDAC complex known to be involved in gene expression and silencing. Part of this activity is carried out through the Tup1 repressor[23]. Tup1 is an interesting example of a repressor that acts as a mediator between transcription factors and chromatin modifiers, having the ability to recruit chromatin modifiers to confer repressed chromatin structure[24]. As Hda1 is one of the chromatin modifiers recruited by Tup1, we were interested in the relation between Tup1 and Yap6. To test whether Yap6 works through Tup1, we examined the Yap6 cohort behavior in a gene expression experiment carried out in a strain deleted for *TUP1* (ref. 19). Indeed, the Yap6 cohort shows a stronger activation in the *tup1Δ* strain experiment (**Fig. 5b**) than in the *hda1Δ* strain experiment, implying that Tup1 participates in the regulation by Yap6 and strongly suggesting that Tup1 is able to recruit additional chromatin modifiers[24]. The high level of overlap between the activated genes in both experiments (hypergeometric $P < 0.003$) (**Supplementary Fig. 2** online) confirms that Tup1 and Hda1 repress these genes by a common mechanism. The common genes are mostly subtelomeric (15/23 genes; $P < 10^{-10}$) and are highly enriched for members of the hexose-transport family (five genes, $P < 0.001$). Thus, our results clearly indicate a role for Yap6 in the regulation of sugar transport that, surprisingly, is affected by Tup1 and Hda1 and not by the chromatin modifiers usually implied in silencing of subtelomeric genes, such as the Sir proteins and Set1 (ref. 25).

Following the Gcn4 example above, we went on to search for chromatin modifiers that affect the Yap6 cohort in a manner opposite to that of Tup1-Hda1. We found that the Yap6 cohort was significantly repressed in a strain with deletion of *SPT3*, a key member of the SAGA complex[26] (**Fig. 5c**). SAGA is a well-characterized HAT complex that acts as a global inducer[27]. Notably, although Spt3 is a SAGA member required for the recruitment of TBP to various SAGA-regulated genes[28], we did not observe any effect on the expression of the Yap6 cohort in mutants with deletion of *GCN5* (SAGA's catalytic subunit) or in strains carrying various mutations in the gene encoding TBP (data not shown). Many of the SAGA complex components can also be found in a different complex called the SAGA-like complex (SILK)[29]. Spt3 was shown to regulate genes through SILK in a manner that does not require SAGA's HAT activity[30]. Thus, our results demonstrate collaboration between Yap6 and Spt3 that is independent of *GCN5*, suggesting the existence of an uncharacterized interactor that provides HAT activity. Analysis of the repressed genes within the Yap6 cohort demonstrates an extensive overlap with genes activated in strains with deletion of *HDA1* ($P < 0.009$) (**Supplementary Fig. 2**). The high overlap between the genes points to an acetylation homeostasis achieved by the Tup1-Hda1 and Spt3-related HAT activities.

TBP-dependent transcription factors

As described above, chromatin modifiers interact with transcription factors to regulate gene expression. The same principle should be applicable to additional proteins that, like the chromatin modifiers, have a wide influence on transcription. TBP, a central activator of transcription, is such a factor. TBP affects transcription of most of the genome and collaborates with cofactors, many of which are chromatin modifiers. Among the TBP cofactors are Mot1 (SWI/SNF-like), Spt3 (HAT), Taf1 (HAT) and the inhibitor NC2 (ref. 31). As TBP is an essential component of the cell, its activity has been modified[32] by overexpressing TBP mutants. The NC2 complex and Taf1 are considered inhibitors of the TBP transcription induction[31]. NC2 acts by competitively inhibiting

Table 3 K-S scores in experiments disrupting various TBP interactions

K-S score	WT	TBPd	DeltaT	NC2	TBPd deltaT
Hap1	–1.18	**–6.83**	1.44	**8.21**	**–6.45**
Skn7	–2.67	**–12.36**	**5.81**	4.15	**–7.47**
Swi4	0.21	**–7.95**	0.73	3.5	**–6.55**

Boldface indicates significant K-S scores.

the TBP association to transcription factor IIA (TFIIA) and TFIIB[32]. Taf1 contains a domain called TANDI that mimics the TATA box and competitively inhibits the TBP interaction with the TATA box[32]. Another TBP inhibition mechanism is through TBP self-dimerization. Mutations have been created[32] that affect TBP dimerization (TBPd), interaction with Taf1 through deletion of the TANDI region (DeltaT) or interaction with NC2 through a mutation in the region encoding the NC2-binding region (NC2). Overexpression of the TBPd mutations leads to a preference in the use of the non-dimerizing mutated TBP. As the TBPd mutations were also shown to have reduced functional capability, we used them in our analysis to search for TBP-dependent transcription factors[32]. We used the other mutants to investigate the regulatory contribution of NC2 and Taf1.

Our results (**Table 3**) support the generally positive regulatory function of the TBP: we observed a clear reduction in gene expression of many cohorts (**Fig. 2b**). Among the TBP-dependent transcription factors, we focused on Hap1, Skn7 and Swi4, three transcription factors that show different interactions with TBP, NC2 and Taf1. For each of these transcription factors, expression level by itself was not sufficient to explain the proposed trend of its cohort (**Supplementary Table 2**).

Hap1 is a transcription factor with roles in the cellular response to heme and oxygen[33]. Its cohort was significantly repressed in a strain carrying the TBPd mutations, indicating that Hap1 is dependent on TBP to induce its genes. Notably, a deletion of the TANDI region of *TAF1* (DeltaT) completely abolished this effect, whereas mutations that affect NC2 binding caused a strong increase in the expression of the Hap1 cohort (**Table 3**). Thus, Hap1 is a good example of a transcription factor that promotes the transcription of its target genes by TBP recruitment but depends on NC2 for their repression.

Skn7 is a transcription factor associated with the oxidative stress response[34]. Like Hap1, its cohort was repressed in strains carrying the TBPd mutations. However, unlike Hap1, the Skn7 cohort also showed a significant induction in the DeltaT strain and, to a lesser extent, also in the strain defective in NC2 interaction (**Table 3**). The less significant effect of Taf1 and NC2 on the Skn7 cohort is likely to be due to a complementary repression by the two mechanisms, each with its own repression targets. Indeed, we observed very little overlap among genes affected by the two regulators (data not shown).

Swi4 is a central cell cycle transcription factor that, together with Swi6, promotes transcription of late G1 genes[35]. The Swi4 cohort was also significantly repressed upon mutation in the region of the gene encoding the TBP dimerization domain, but unlike Hap1 and Skn7, its cohort depended neither on Taf1 nor on the NC2 repressor (**Table 3**). Thus, in the case of Swi4, if there is a repression mechanism that works through the TBP, it is conferred by factors other than the ones tested here (Taf1, NC2). In all the experiments carried out in strains lacking both TBP dimerization and the TANDI region (DeltaT TBPd), we observed a strong reduction of expression, similar to the one seen in strains affected for dimerization only (**Table 3**). This epistatic effect of the TBP destabilizing mutation points to a need for a functional TBP in the Taf1-mediated regulation.

As mentioned above, TBP regulates gene expression by binding AT-rich sequences called TATA boxes. TATA box–containing genes comprise ~20% of the yeast genome[36]. An analysis of the distribution of TATA box occupancy (**Supplementary Table 4** online) showed that, as expected, the TBP-dependent cohorts were highly enriched for TATA box–containing genes (Hap1, Skn7 and Swi4 cohorts had hypergeometric $P < 10^{-14}$, $P < 10^{-20}$ and $P < 10^{-9}$, respectively). Thus, our analysis shows that the TBP is central in the regulation carried out by several transcription factors.

Conclusions

Chromatin organization has a central role in many biological mechanisms and particularly in transcription. Although many factors have been found to participate in the regulation of the chromatin structure, thus far there has not been a systematic study of their global contribution to transcription. In this work, using a compendium of genome-wide profiles of strains defective in chromatin modifier activity, we lay the infrastructure to the study of chromatin modifier–transcription factor interactions and their relation to transcription. We show that this approach is able to detect cooperation between a transcription factor and chromatin modifiers even when a complex combinatorial regulation is involved. Our systematic analysis of all available transcription factor cohorts against the large gene expression compendium provides the first global picture in a eukaryote of the complex regulation by transcription factors in the context of chromatin organization. We have shown that our method is robust enough to detect novel regulation mechanisms of well-characterized transcription factors (for example, Ume6 and Gcn4) as well as exposing features of the regulation of uncharacterized transcription factors, such as Yap6. Note that our test cannot distinguish between direct and indirect chromatin modifier–transcription factor interaction. The difficulty in separating direct effects from indirect effects is prevalent in many studies on gene regulation networks[37–40].

The majority of the profiles in the compendium were measured under standard growth conditions (for example, rich medium). By using our method on transcriptional profiles obtained in other environments, we can start to investigate the mechanism by which environmental conditions lead to differential gene expression. In addition, as many chromatin modifiers are evolutionarily conserved[41], the generation of similar experiments in higher eukaryotes will provide insights into the evolution of global regulatory mechanisms.

METHODS

Yeast genome. We retrieved 6,646 yeast ORFs from the *Saccharomyces* Genome Database (http://www.yeastgenome.org) (July 2005). To avoid cross-hybridization biases in the gene expression and location data set, 103 ORFs containing mitochondrial genes and short dubious ORFs were ignored in the analysis.

Data preparation. From 26 publications, we collected 170 gene expression profiles obtained with strains mutated for various chromatin modifiers. The complete list of publications and experiments is available in **Supplementary Table 1**. Data were downloaded from papers' web supplements. Normalization was done as in ref. 42.

Transcription factor–DNA binding profiles were obtained from ref. 1. A P value cutoff of 0.001 was used to define the set of genes bound by a particular transcription factor (the transcription factor cohort).

To account for the strong correlated response of the ribosomal genes[42] in most experiments, all transcription factors that were found to be significantly enriched ($P < 0.001$) in ribosome-related GO terms were excluded from the analysis. Our analysis used the remaining 75 cohorts, containing at least 50 genes, that originated from 49 transcription factors tested in different environments.

Kolmogorov-Smirnov Test. Given two samples of values, the Kolmogorov-Smirnov (K-S) test[8] is designed to examine whether they have the same value distribution. The main advantage of this test is that it makes no assumption on the distributions from which the samples originated. This is important when dealing with expression profiles from different sources.

This test is appropriate for two reasons. First, owing to its nonparametric nature, the test is robust and suitable for heterogeneous data, such as our diverse expression profiles, that originate from many studies. Second, the test also provides a P value for the statistical significance of the difference between the two distributions.

For each value v, the K-S test measures the difference (between the control and the cohort samples) in the fraction of genes that have an expression value lower than v. The K-S statistic is defined to be the maximum absolute value of that difference.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

In the case of the null hypothesis (that the two samples originate from the same distribution), the distribution of the statistic can be calculated, and a *P* value ($P_{K\text{-}S}$) can be assigned to the disparity between the two samples[8].

The K-S score is defined as K-S = $-\log_{10}(P_{K\text{-}S})$ if the difference realizing the statistic is positive and $\log_{10}(P_{K\text{-}S})$ otherwise. Hence, the absolute value of the K-S score indicates significance of the disparity, and its sign indicates the direction of the disparity: a positive sign shows that the cohort genes tend to have higher values than the rest of the genes. A significant disparity is defined as scores with absolute value above 5.41 (Bonferroni corrected *P* value < 0.05).

Altered gene groups and their overlap test. A gene is considered altered in a gene expression experiment if its *Z* score is >1. Given a gene expression experiment *E* with average μ and s.d. σ and a transcription factor cohort S (the transcription factor target gene group), the elevated cohort genes are defined as TF_E = {g in S| E(g) > μ + σ } and the set of genes with reduced expression is defined as {g in S| E(g) < μ – σ }.

Given two altered (elevated or reduced) subgroups of S, S_1 and S_2, the significance of their overlap is calculated using the hypergeometric distribution, where S is considered as the sample pool.

Annotation enrichment. All GO annotations were taken from the Gene Ontology database[43] (July 2005). Annotation enrichments were calculated using TANGO[44].

Hierarchical clustering. Hierarchical clustering of the cohorts and the experimental conditions based on the significant K-S scores matrix (all |K-S scores| > 1.3) was carried out using the EXPANDER analysis and visualization tool (version 3.0)[44].

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS
We thank A. Tanay for guidance and ideas and all the members of the Kupiec and Shamir laboratories for discussions and encouragement. This work was supported in part by grants from the Israel Science Foundation to M.K. and R.S. and a grant from the Israeli Ministry of Science and Technology to M.K.

COMPETING INTERESTS STATEMENT
The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

1. Harbison, C.T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99–104 (2004).
2. Kurdistani, S.K. & Grunstein, M. Histone acetylation and deacetylation in yeast. *Nat. Rev. Mol. Cell Biol.* **4**, 276–284 (2003).
3. Tsukiyama, T. The *in vivo* functions of ATP-dependent chromatin-remodelling factors. *Nat. Rev. Mol. Cell Biol.* **3**, 422–429 (2002).
4. Eberharter, A. & Becker, P.B. Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. *EMBO Rep.* **3**, 224–229 (2002).
5. Kadosh, D. & Struhl, K. Repression by Ume6 involves recruitment of a complex containing Sin3 corepressor and Rpd3 histone deacetylase to target promoters. *Cell* **89**, 365–371 (1997).
6. Swanson, M.J. *et al.* A multiplicity of coactivators is required by Gcn4p at individual promoters *in vivo*. *Mol. Cell. Biol.* **23**, 2800–2820 (2003).
7. Lee, T.I. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae. Science* **298**, 799–804 (2002).
8. Stephens, M.A. Use of the Kolmogorov-Smirnov, Cramer-von Mises and related statistics without extensive tables. *J. R. Stat. Soc. [Ser A]* **32**, 115–122 (1970).
9. Fazzio, T.G. *et al.* Widespread collaboration of Isw2 and Sin3-Rpd3 chromatin remodeling complexes in transcriptional repression. *Mol. Cell. Biol.* **21**, 6450–6460 (2001).
10. Goldmark, J.P., Fazzio, T.G., Estep, P.W., Church, G.M. & Tsukiyama, T. The Isw2 chromatin remodeling complex represses early meiotic genes upon recruitment by Ume6p. *Cell* **103**, 423–433 (2000).
11. Robyr, D. *et al.* Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. *Cell* **109**, 437–446 (2002).
12. Rubin-Bejerano, I., Mandel, S., Robzyk, K. & Kassir, Y. Induction of meiosis in *Saccharomyces cerevisiae* depends on conversion of the transcriptional represssor Ume6 to a positive regulator by its regulated association with the transcriptional activator Ime1. *Mol. Cell. Biol.* **16**, 2518–2526 (1996).
13. Sabet, N., Volo, S., Yu, C., Madigan, J.P. & Morse, R.H. Genome-wide analysis of the relationship between transcriptional regulation by Rpd3p and the histone H3 and H4 amino termini in budding yeast. *Mol. Cell. Biol.* **24**, 8823–8833 (2004).
14. Clapier, C.R., Langst, G., Corona, D.F., Becker, P.B. & Nightingale, K.P. Critical role for the histone H4 N terminus in nucleosome remodeling by ISWI. *Mol. Cell. Biol.* **21**, 875–883 (2001).
15. Fazzio, T.G., Gelbart, M.E. & Tsukiyama, T. Two distinct mechanisms of chromatin interaction by the Isw2 chromatin remodeling complex *in vivo*. *Mol. Cell. Biol.* **25**, 9165–9174 (2005).
16. Hinnebusch, A.G. Translational regulation of GCN4 and the general amino acid control of yeast. *Annu. Rev. Microbiol.* **59**, 407–450 (2005).
17. Kuo, M.H., vom Baur, E., Struhl, K. & Allis, C.D. Gcn4 activator targets Gcn5 histone acetyltransferase to specific promoters independently of transcription. *Mol. Cell* **6**, 1309–1320 (2000).
18. Natarajan, K., Jackson, B.M., Zhou, H., Winston, F. & Hinnebusch, A.G. Transcriptional activation by Gcn4p involves independent interactions with the SWI/SNF complex and the SRB/mediator. *Mol. Cell* **4**, 657–664 (1999).
19. Hughes, T.R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
20. Natarajan, K. *et al.* Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. *Mol. Cell. Biol.* **21**, 4347–4368 (2001).
21. Keogh, M.C. *et al.* Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123**, 593–605 (2005).
22. Mendizabal, I., Rios, G., Mulet, J.M., Serrano, R. & de Larrinoa, I.F. Yeast putative transcription factors involved in salt tolerance. *FEBS Lett.* **425**, 323–328 (1998).
23. Wu, J., Suka, N., Carlson, M. & Grunstein, M. TUP1 utilizes histone H3/H2B-specific HDA1 deacetylase to repress gene activity in yeast. *Mol. Cell* **7**, 117–126 (2001).
24. Smith, R.L. & Johnson, A.D. Turning genes off by Ssn6-Tup1: a conserved system of transcriptional repression in eukaryotes. *Trends Biochem. Sci.* **25**, 325–330 (2000).
25. Santos-Rosa, H., Bannister, A.J., Dehe, P.M., Geli, V. & Kouzarides, T. Methylation of H3 lysine 4 at euchromatin promotes Sir3p association with heterochromatin. *J. Biol. Chem.* **279**, 47506–47512 (2004).
26. Grant, P.A. *et al.* Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes Dev.* **11**, 1640–1650 (1997).
27. Huisinga, K.L. & Pugh, B.F. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *Saccharomyces cerevisiae*. *Mol. Cell* **13**, 573–585 (2004).
28. Bhaumik, S.R. & Green, M.R. Differential requirement of SAGA components for recruitment of TATA-box-binding protein to promoters *in vivo*. *Mol. Cell. Biol.* **22**, 7365–7371 (2002).
29. Sterner, D.E., Belotserkovskaya, R. & Berger, S.L. SALSA, a variant of yeast SAGA, contains truncated Spt7, which correlates with activated transcription. *Proc. Natl. Acad. Sci. USA* **99**, 11622–11627 (2002).
30. Belotserkovskaya, R. *et al.* Inhibition of TATA-binding protein function by SAGA subunits Spt3 and Spt8 at Gcn4-activated promoters. *Mol. Cell. Biol.* **20**, 634–647 (2000).
31. Pugh, B.F. Control of gene expression through regulation of the TATA-binding protein. *Gene* **255**, 1–14 (2000).
32. Chitikila, C., Huisinga, K.L., Irvin, J.D., Basehoar, A.D. & Pugh, B.F. Interplay of TBP inhibitors in global transcriptional control. *Mol. Cell* **10**, 871–882 (2002).
33. Zhang, L. & Hach, A. Molecular mechanism of heme signaling in yeast: the transcriptional activator Hap1 serves as the key mediator. *Cell. Mol. Life Sci.* **56**, 415–426 (1999).
34. Raitt, D.C. *et al.* The Skn7 response regulator of *Saccharomyces cerevisiae* interacts with Hsf1 *in vivo* and is required for the induction of heat shock genes by oxidative stress. *Mol. Biol. Cell* **11**, 2335–2347 (2000).
35. Bahler, J. Cell-cycle control of gene expression in budding and fission yeast. *Annu. Rev. Genet.* **39**, 69–94 (2005).
36. Basehoar, A.D., Zanton, S.J. & Pugh, B.F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**, 699–709 (2004).
37. Hartemink, A.J., Gifford, D.K., Jaakkola, T.S. & Young, R.A. Combining location and expression data for principled discovery of genetic regulatory network models. *Pac. Symp. Biocomput.*, 437–449 (2002).
38. Gardner, T.S., di Bernardo, D., Lorenz, D. & Collins, J.J. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* **301**, 102–105 (2003).
39. Segal, E. *et al.* Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.* **34**, 166–176 (2003).
40. Workman, C.T. *et al.* A systems approach to mapping DNA damage response pathways. *Science* **312**, 1054–1059 (2006).
41. Khorasanizadeh, S. The nucleosome: from genomic organization to genomic regulation. *Cell* **116**, 259–272 (2004).
42. Tanay, A., Steinfeld, I., Kupiec, M. & Shamir, R. Integrative analysis of genome-wide experiments in the context of a large high-throughput data compendium. *Mol. Syst. Biol.* **2005**, 2005.0002 (2005).
43. Harris, M.A. *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, D258–D261 (2004).
44. Shamir, R. *et al.* EXPANDER–an integrative program suite for microarray data analysis. *BMC Bioinformatics* **6**, 232 (2005).
45. Bernstein, B.E. *et al.* Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc. Natl. Acad. Sci. USA* **99**, 8695–8700 (2002).
46. Bernstein, B.E., Tong, J.K. & Schreiber, S.L. Genomewide studies of histone deacetylase function in yeast. *Proc. Natl. Acad. Sci. USA* **97**, 13708–13713 (2000).