Nuclear colocalization of transcription factor target genes strengthens coregulation in yeast

Zhiming Dai* and Xianhua Dai*

Department of Electronics, School of Information Science and Technology, Sun Yat-Sen University, Guangzhou 510006, China

ABSTRACT

Eukaryotic chromosomes are not randomly distributed in the interphase nucleus, but instead occupy distinct territories. Nonetheless, the genome-wide relationships of gene regulation to gene nuclear location remain poorly understood in yeast. In the three-dimensional view of gene regulation, we found that a considerable number of transcription factors (TFs) regulate genes that are colocalized in the nucleus. Colocalized TF target genes are more strongly coregulated compared with the other TF target genes. Target genes of chromatin regulators are also colocalized. These results demonstrate that colocalization of coregulated genes is a common process, and three-dimensional gene positioning is an important part of gene regulation. Our findings will have implications in understanding nuclear architecture and function.

INTRODUCTION

The three-dimensional organization of chromosomes is not random within the eukaryotic nucleus (1–5). Individual chromosomes are not randomly scattered over the whole nuclear space but occupy distinct positions termed chromosome territories (6,7). Gene-rich chromosomes tend to occupy interior positions in the nucleus, whereas, gene-poor chromosomes tend to be peripherally located in human (8). The spatial conformation of chromosomes results in interchromosomal interactions that bring distant chromosomes into close proximity (9–12).

The non-random spatial features of chromosomes are linked to gene regulation. There is increasing evidence that much mammalian transcription occurs within a limited number of discrete nuclear structures known as transcription factories (13,14). Experiments on individual genes have demonstrated that genes on different chromosomes might migrate to the transcription factory, leading to interchromosomal interactions (15,16). When transcription initiation is inhibited, active genes disassociate from transcription factories (17). This result demonstrates that transcriptional regulation might tether distal active genes to common loci to form interchromosomal interactions. Although several coregulated galactose *GAL* genes in budding yeast were colocalized in the nucleus (18), whether or not transcription factories exist in budding yeast has not been clearly delineated. A recent study has found that tRNA genes and early replication origin genes in yeast are enriched in interchromosomal interactions, but other specific categories of genes, such as cell cycle-regulated genes, Pol II-regulated genes, DNA repair genes, etc., are not enriched in interchromosomal interactions among their members (19).

Transcription factors (TFs) control gene expression by binding regulatory regions of relevant genes. Their target genes are not randomly distributed across linear chromosomes. The majority of TFs show a strong preference to regulate genes on specific chromosomes, and their target genes are also positionally clustered on a chromosome (20,21). These results indicate that transcriptional regulation constrains organization of genes across and on the chromosomes. However, whether TF target genes are randomly distributed in the nucleus or encoded on specific regions remains to be elucidated.

It has become clear that the traditional one-dimensional view of gene regulation is oversimplified. Despite the relationship of gene regulation with spatial organization of chromosomes observed on individual genes, it is still not clear how gene regulation influences gene nuclear location on a genomic scale. Although a previous study has mapped the three-dimensional architecture of whole human genome (22), the data were in low resolution and thus, there is insufficient data in the human system for the investigation into how gene regulation influences gene nuclear location on a genomic scale. Until recently, three-dimensional maps of the yeast genome were generated with high resolution (19,23). These valuable data allow for a direct genome-wide examination of gene nuclear location and its relationship with gene regulation. Analyzing these data (19), we found that target genes of a considerable number of TFs

*To whom correspondence should be addressed. Tel: +86 020 3994 3331; Fax: +86 020 3994 3331; Email: zhimdai@gmail.com
*Correspondence may also be addressed to X. Dai. Email: issdxh@mail.sysu.edu.cn

tend to be colocalized in the nucleus. This colocalization results in tighter coregulation of TF target genes. We also found that target genes of other *trans* regulators, including chromatin remodelers and histone modified enzymes, tend to be colocalized. Genes with high *trans* effects on gene expression divergence are also clustered together in the nucleus.

MATERIALS AND METHODS

Data Preparation

Genome-wide changes in gene expression data corresponding to the knockout of 269 TFs were taken from Hu *et al.* (24). As a TF can regulate secondary targets via regulatory cascades, we used the refined transcriptional regulatory network that eliminated indirect regulatory interactions (24). We determined knockout target genes for each TF. As we used hypergeometric distribution to evaluate nuclear colocalization of TF target genes, TFs with fewer target genes could bias the statistical reliability of the computed *P*-values. In this study, we restricted the analysis to TFs with more than 20 knockout target genes, resulting in a total of 174 TFs. We also separated activated target genes from repressed target genes for each of the 174 TFs. Genome-wide binding data corresponding to 203 TFs were taken from Harbison *et al.* (25). We determined binding target genes for each TF. We also restricted the analysis to TFs with more than 20 binding target genes, resulting in a total of 158 TFs. A *P*-value cutoff of 0.005 was used to define the set of genes affected by the knockout of a particular TF or bound by a particular TF. We chose this relaxed criterion of confidence (*P*-value) in order to increase the number of true and functional TF–DNA interactions that can be found. We also repeated the analysis using a stricter *P*-value cutoff of 0.001 as that in the original literatures (24,25), and found that the resulting statistical significance (*P*-values) of colocalization are significantly correlated with those for the *P*-value cutoff of 0.005 (Pearson correlation coefficient, $R = 0.61$, $P < 10^{-28}$ for TF knockout target genes, $R = 0.52$, $P < 10^{-20}$ for TF binding target genes).

Genome-wide interchromosomal and intrachromosomal interaction data were taken from Duan *et al.* (19). These data represent a time- and population-averaged snapshot of events in the cell. As in the original literature (19), a false discovery rate (FDR) value cutoff of 0.01 was used to define interchromosomal and intrachromosomal interactions. 240 629 interchromosomal interactions and 65 683 intrachromosomal interactions among 3991 segments on different chromosomes, with kilobase resolution, have been identified. Of all intrachromosomal interactions, ~20% are between segments separated by <60 kb. In the original literature (19), intrachromosomal interactions that are between segments separated by <20 kb were eliminated. In this study, we set a stricter cutoff to control for the interactions caused by linear proximity in the genome. We eliminated intrachromosomal interactions in which the two segments are separated by <60 kb. Two segments on different chromosomes that interact with each other are colocalized in the nucleus. For each segment, we calculated its distance to transcription start sites (TSS) of all genes on its corresponding chromosome, and identified the gene whose TSS is the closest to it. Every segment has its corresponding gene. In this way, we could identify gene pairs that are colocalized in the nucleus according to segment pairs that are colocalized (i.e. show interchromosomal interactions). In this study, we only used this strategy to calculate the pair-wise Pearson expression correlation coefficient among all gene pairs that are colocalized in the nucleus to evaluate whether colocalized gene pairs, regardless of coregulation or not, are more coexpressed than random gene pairs.

The list of TFs, which show a preference to regulate genes that are positionally clustered on a chromosome, was taken from Janga *et al.* (20). We evaluated the overlap between target-colocalized TFs and this list of TFs. Protein abundance data was taken from Newman *et al.* (26), which was measured by high-throughput flow cytometry coupled with green fluorescent protein (GFP)-tagged yeast strains at single-cell resolution. This data set includes the number of molecules for more than 2500 proteins in yeast. We compared protein abundance values between target-colocalized TFs and the other TFs. Genome-wide gene transcription activity (transcription rate) data were taken from Holstege *et al.* (27). The TSS data was taken from David *et al.* (28).

Genome-wide gene expression data used for coexpression analysis were measured under normal growth conditions (29–31), a total of 112 time points. Histone modification data were taken from ChromatinDB (32), a database of genome-wide histone modification patterns for *Saccharomyces cerevisiae*. We added the histone modification data from Pokholok *et al.* (33), resulting in a total of 25 histone modifications. For each promoter (500-bp upstream of the gene in this study, the upstream region was truncated if it overlapped with neighboring genes), we calculated the average level for each histone modification, respectively. For each histone modification, we calculated the Pearson correlation coefficient between average modification levels at promoters and gene transcription activity levels. We identified cohort of genes that show high modification levels at promoters for each modification. Genes belong to one modification cohort if they display significantly high levels (Z-score > 1.64, $P < 0.05$) of the corresponding modification.

Genome-wide chromatin remodeler occupancy data at TSS and upstream activation sequence (UAS) were taken from Venters *et al.* (34). We identified cohorts of genes that show high occupancy levels at UAS and TSS for each remodeler. Genes belong to one remodeler cohort if they display significantly high occupancy levels (Z-score > 1.64, $P < 0.05$) of the corresponding remodeler. For each remodeler, we calculated the Pearson correlation coefficient between occupancy levels and gene transcription activity levels. We used a compendium of gene expression experiments in which various chromatin modifiers were deleted or mutated (35). Genes belong to one remodeler cohort if they display significant changes ($|Z - \text{score}| > 1.64$, $P < 0.05$) in gene expression accompanying the perturbation of the corresponding remodeler.

Genome-wide RNA polymerase II occupancy (RNA polymerase II subunit Rpo21) data were taken from Venters *et al.* (34). For each promoter, we calculated the average RNA polymerase II occupancy level. We identified cohort of genes that show high RNA polymerase II occupancy at promoters. Genes belong to RNA polymerase II-enriched promoters if they display significantly high RNA polymerase II occupancy levels (Z-score > 1.64, $P < 0.05$). The list of occupied proximal nucleosome (OPN) genes was taken from Tirosh and Barkai (36). H2A.Z nucleosome data was taken from Albert *et al.* (37). To avoid confusion, we restricted the analysis to the 5% most scored H2A.Z nucleosomes in the literature. We identified genes that contain H2A.Z nucleosomes in promoter regions. Data for *cis* and *trans* effects on gene expression divergence between species were taken from Tirosh *et al.* (38). The relative contribution of *cis* and *trans* effects to gene expression divergence has been experimentally measured by using the hybrid of *S. cerevisiae* and *Saccharomyces paradoxus*. As both alleles of each gene are under the same nuclear environment (the same *trans* effects) in the hybrid, differences in their expression reflect *cis* effects on expression divergence, whereas expression differences between the parental genes that disappear in the hybrid reflect *trans* effects. We defined the genes whose both alleles show significant difference in gene expression ($|Z - \text{score}| > 1.64$, $P < 0.05$) within the hybrid as genes with high *cis* effects. The same threshold ($|Z - \text{score}| > 1.64$, $P < 0.05$) was used to define genes with high *trans* effects.

Statistical method

Given two samples of values, the Mann–Whitney U-test is designed to examine whether they have equal medians. The main advantage of this test is that it makes no assumption that the samples are from normal distributions.

Evaluation of colocalization

We evaluated whether a group of genes is colocalized in the nucleus. Considering yeast interchromosomal interaction data for an example, we identified the segments that are close to each gene [within 2.5 kb to the TSS of gene. This cutoff value was chosen according to the original literature (19)] in yeast, and assigned each gene with its identified segments. The distribution of segments might be non-uniform in the genome. Some genes might have more segments close to them than other genes. If one segment A shows interchromosomal interaction with another segment B, it is more likely that its adjacent segment C also shows interaction with the segment B. If segments A and C are assigned to one gene, it will inevitably lead to an overall elevated number of interactions for this gene. We next evaluated whether two genes have interchromosomal interactions or not. For two given genes, if there is at least one pair of their segments showing interchromosomal interactions, these two genes were defined as having interchromosomal interactions. Note that, we only considered whether two genes have interchromosomal interactions or not, regardless of how many pairs of their segments show interchromosomal

interactions. This criterion controlled for some genes that have more assigned segments. We calculated the number M of all possible interchromosomal interactions among all 5761 genes ($M = 15\,335\,972$), and counted the actual number K of all experimentally identified interchromosomal interactions ($K = 1\,163\,762$). These numbers were considered as the background level (genome-wide level). K/M was the percent of genome-wide level. Given a group of genes, we calculated the number N of all possible interchromosomal interactions among these genes, and counted the actual number X of experimentally identified interchromosomal interactions among these genes. X/N was the percent of this gene group. Given a group of genes, to further control for linear proximity in the genome, we only considered genes that are separated by at least 60 kb. The probability of randomly drawing exactly X interactions of K in N drawings without replacement from M is given by a hypergeometric distribution:

$$P(X = x) = \frac{\binom{N}{x}\binom{M - N}{K - x}}{\binom{M}{K}}.$$

Then, the *P*-value is:

$$P-\text{value} = 1 - \sum_{i=0}^{X-1} P(X = i).$$

We used the resulting *P*-value to evaluate whether this group of genes is colocalized in the nucleus as in a previous study (19).

We also compared the enrichment of interchromosomal interactions for a group of genes to background levels on individual chromosomes. We calculated for all genes on one specific chromosome, the number M of all their possible interchromosomal interactions with all genes on the other chromosomes, and counted the actual number K of all experimentally identified interchromosomal interactions. These numbers were considered as the background level. Given a group of genes, we calculated for the genes on the specific chromosome, the number N of all their possible interchromosomal interactions with the other genes on the other chromosomes, and counted the actual number X of experimentaly identified interchromosomal interactions.

RESULTS

A significant number of TFs regulate genes that are colocalized in the nucleus

To investigate the relationship between transcriptional regulation and gene location in the nucleus, we used two genome-wide yeast data sets. One is target genes of various TFs. If one gene is the target gene of one TF, it should be differentially expressed after this TF is knocked out. We determined target genes for each TF using genome-wide gene expression data corresponding to the knockout of TFs (24). The other data set is the

genome-wide interchromosomal and intrachromosomal interactions data (19), which reflects the topologies and spatial relationships of chromosomes. Chromosomal interaction partners have been determined with kilobase resolution for nearly 4000 segments on different chromosomes, by coupling chromosome conformation capture-on-chip (4C) and massively parallel sequencing. There is an inverse relationship between intrachromosomal interaction frequency and genomic distance separating interacting segments (Supplementary Figure S1).

We examined whether TF target genes are colocalized in the nucleus. We restricted the analysis to TFs with more than 20 target genes. We found that 34 of 174 TFs ($\sim$20%) have target genes that are colocalized (hypergeometric $P < 0.01$, after Bonferroni correction for multiple testing; see 'Materials and Methods' section; Supplementary Table S1). When including TFs with fewer than 20 target genes, the number of TFs whose target genes are colocalized remained the same. As different chromosomes in yeast can have fundamentally different properties in terms of chromosomal interactions (19), we asked whether TF target genes show enrichment of interchromosomal interactions on specific chromosomes. We compared the enrichment of interchromosomal interactions for TF target genes to background levels on individual chromosomes using hypergeometric test. We found that target genes of the 34 TFs show enrichment of interchromosomal interactions on specific chromosomes, respectively (Supplementary Table S2). We calculated pair-wise linear distances among target genes for each TF, and found that the resulting distances of the 34 TFs are comparable with those of the other TFs (303 798 bp versus 304 689 bp; $P = 1$, Mann–Whitney U-test). To further control for linear proximity in the genome, we repeated the analysis only considering TF target genes that are separated by at least 60 kb. This strict criterion filtered out 10 TFs, resulting in 24 TFs whose target genes are colocalized (hypergeometric $P < 0.01$, after Bonferroni correction; Figure 1 and Supplementary Table S1). These 24 TFs are hereinafter referred to as target-colocalized TFs. In the following analysis, we used the strict criterion to examine whether genes are enriched in interchromosomal interactions.

We next examined the nuclear colocalization of TF target genes using intrachromosomal interactions data. We found that 1 of 174 TFs have target genes that are significantly enriched in intrachromosomal interactions (hypergeometric, $P < 0.01$, after Bonferroni correction; Supplementary Figure S2). Moreover, target genes of all TFs show significantly positive correlation between their enrichment in interchromosomal interactions and that in intrachromosomal interactions (Pearson correlation coefficient, $R = 0.51$, $P < 10^{-12}$; Supplementary Figure S3). We repeated the analysis for activated and repressed target gene cohorts separately and found similar observations regardless of the direction of the expression change (Supplementary Figure S4). To strictly control for linear proximity in the genome, we used interchromosomal interactions in the following analysis. Two segments on different chromosomes that interact with each other are supposed to be colocalized in the nucleus. In the following

analysis, the colocalization in the nucleus represents the enrichment of interchromosomal interactions.

Genome-wide analysis has revealed that TF binding genes detected by chromatin immunoprecipitation (ChIP)-chip (25) show a low overlap with TF knockout target genes. We thus examined the relationship between transcriptional regulation and gene nuclear location using large-scale TF binding data measured by ChIP-chip (25). We found that 5 of 158 TFs have binding target genes that are colocalized (hypergeometric, $P < 0.01$, after Bonferroni correction; Supplementary Figure S5). We collected high-quality ChIP-chip data sets for TFs Swi4 (39), Hsf1 (40) and Rap1 (41). The binding target genes of these three TFs are not colocalized according to the large-scale ChIP-chip data sets (25). However, the binding target genes of two among these three TFs are colocalized according to the high-quality ChIP-chip data sets (hypergeometric, $P < 10^{-4}$ for Swi4; $P < 10^{-17}$ for Hsf1; $P = 1$ for Rap1), suggesting that data quality may be one reason for the low colocalization. TF binding detected by ChIP-chip is likely to be confounded by stochastic binding and may vary depending on the physiology of the cells at the experimental time. In addition, most knockout target genes are enriched in known TF binding DNA motifs in the corresponding promoters (24). In the following analysis, we used TF knockout target data sets instead of TF binding target data sets.

We next investigated the relationship between the distributions of TF target genes in the nucleus and those on linear chromosomes. A previous study has analyzed the distribution of TF target genes on chromosomes, and has revealed that most TFs show a preference to regulate genes that are positionally clustered on a chromosome (20). We asked whether TFs whose target genes are colocalized in the nucleus also tend to regulate genes clustered together on a chromosome. Using the list of TFs that tend to regulate genes positionally clustered on a chromosome (20), we found that target-colocalized TFs are not significantly enriched in, or depleted of TFs that tend to regulate genes positionally clustered on a chromosome (hypergeometric, $P = 0.29$ for enrichment; hypergeometric, $P = 0.88$ for depletion). This result suggests that there is no general relationship between positional preferences on linear chromosomes and those in the nucleus for TF target genes.

Colocalization of TF target genes strengthens coregulation

We examined whether colocalization of TF target genes has effect on TF regulation. The colocalization of TF target genes should facilitate TF regulation compared to the scattered distribution of TF target genes. The function of TF binding to regulatory regions is to control gene expression. Genes coregulated by a common TF are expected to be coexpressed (42). The degree of expression coherence for TF target genes signifies the degree of TF regulation (43). We asked whether colocalization of TF target genes strengthens the degree of TF regulation. We used a combined gene expression data set on 112 time points during the condition (29–31) similar to that where TF knockout data and three-dimensional data were measured.
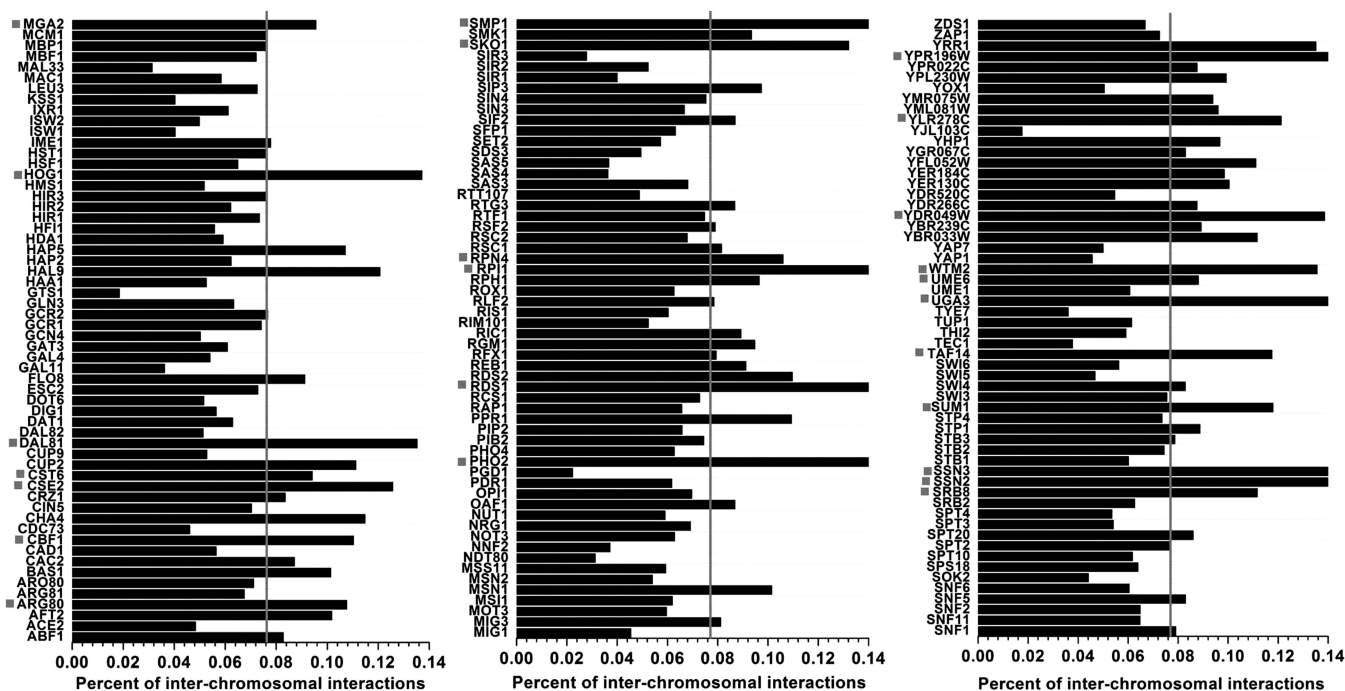
Figure 1. Enrichment of interchromosomal interactions with regard to 174 groups of TF target genes (TF knockout data). Each bar represents the enrichment or depletion (the ratio between actual number and possible number) relative to the percent (gray line) of genome-wide level. Results are shown for strict criterion, i.e. exclusion of TF target genes that are separated by <60 kb. See Supplementary Table S1 for results of both criteria and the full list of hypergeometric *P*-values. TFs with a gray square on the left are significantly enriched (hypergeometric, $P < 0.01$, after Bonferroni correction for multiple testing).
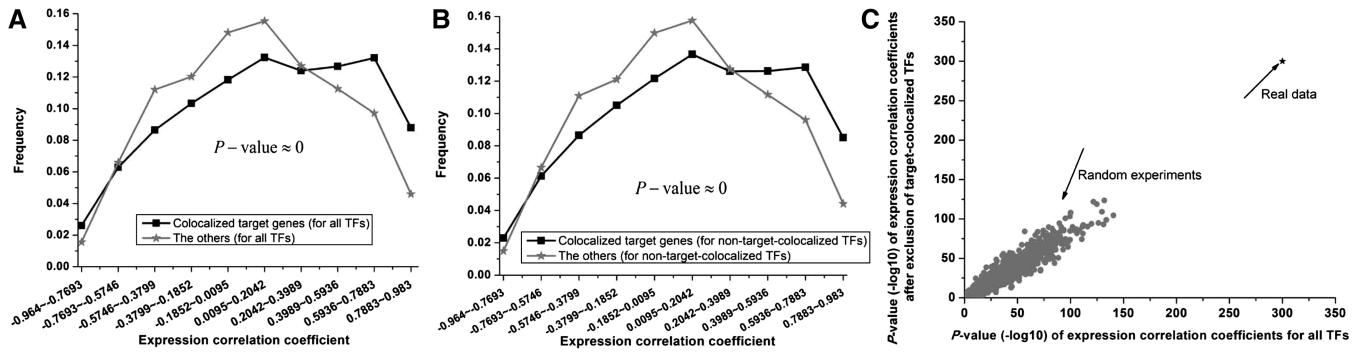
We separated target genes that are colocalized (i.e. show interchromosomal interactions) from those that are not for each TF. We calculated for each TF the pair-wise expression correlation coefficient among the two separated gene groups, respectively. To control for linear proximity, we excluded pair-wise Pearson expression correlation coefficient between gene pairs that are on the same chromosome. Colocalized target genes are more coexpressed than the other target genes (correlation coefficient of 0.14 versus 0.06; $P \approx 0$, Mann–Whitney U-test; Figure 2A). When excluding target-colocalized TFs for analysis, similar observation could be reproduced (correlation coefficient of 0.14 versus 0.06; $P \approx 0$, Mann–Whitney U-test; Figure 2B). We asked whether this is a general feature of all TFs. We found that only five among all TFs, including Rsc1, Tec1, Sas3, Gts1 and Cad1, violate this feature. Their colocalized target genes are less coexpressed than the other target genes ($P < 0.01$, Mann–Whitney U-test, after Bonferroni correction). To examine the artifact caused by the difference in the sizes between colocalized target genes and the other target genes, we used random perturbation experiments to test the validity of the observations above. For every TF, fixing the number of target genes that are colocalized, we perturbed target genes that are colocalized with those that are not at random from its target gene cohort. Using this new randomized map, we repeated the analysis as above to evaluate the statistical difference of pair-wise expression correlation coefficients between target genes that are colocalized and those that are not. We repeated this randomized experiment 1000 times. We found that the statistical differences between

target genes that are colocalized and those that are not for these randomized data are weaker than those of the real data (Figure 2C).

However, the apparent correspondence between colocalization of TF target genes and high degree of coregulation might be caused by the intrinsic colocalization of target genes, not by their common regulatory systems. For example, neighboring gene pairs on linear chromosomes tend to be coexpressed (44), but this coexpression can not be attributed to their shared TF regulation (45). We examined whether the correspondence is caused by the intrinsic colocalization of TF target genes. If this is the case, colocalized gene pairs, regardless of coregulation or not, should be more coexpressed than random gene pairs. We calculated the pair-wise Pearson expression correlation coefficient among all gene pairs that are colocalized in the nucleus. We also repeated the calculation for all possible gene pairs. There is no statistical difference in degree of coexpression between these two types of gene pairs (correlation coefficient of 0.02 versus 0.02; $P = 0.08$, Mann–Whitney U-test; Supplementary Figure S6). This result excludes the possibility that the tight coregulation of TF colocalized target genes is caused by their intrinsic colocalization.

Chromatin regulatory genes are colocalized in the nucleus

Having revealed the colocalization of TF target genes in the nucleus, one natural question is whether target genes of other *trans* regulators are also colocalized. In addition to TF regulation, another important gene regulatory mechanism is at the chromatin level. Chromatin

Figure 2. Colocalization of TF target genes corresponds to strong coregulation. (**A**) We separated target genes that are colocalized from those that are not for each TF. We calculated for each TF the pair-wise expression correlation coefficient among the two separated gene groups, respectively. Distributions of resulting correlation coefficient values are presented for the two gene groups. Higher positive correlation indicates strong coregulation. Colocalized target genes are more strongly coregulated than the other target genes ($P \approx 0$, Mann–Whitney U-test). (**B**) Same as (A), but exclusion of target-colocalized TFs for analysis. Colocalized target genes are still more strongly coregulated than the other target genes ($P \approx 0$, Mann–Whitney U-test). (**C**) Scatter comparison of *P*-values (Mann–Whitney U-test, $-$log10 transformed) that evaluate the difference of pair-wise expression correlation coefficients between target genes that are colocalized and those that are not. *X*-axis is the *P*-values for all TFs, and *y*-axis is the *P*-values after exclusion of target-colocalized TFs for analysis. The black star is the *P*-values for real data [(A) and (B)]. The gray dots are the *P*-values for the 1000 random perturbation experiments. To control for linear proximity, we excluded pair-wise Pearson expression correlation coefficient between gene pairs that are on the same chromosome in (A–C).

regulators influence gene expression levels by controlling chromatin structure. In general, there are two main ways in which cells regulate chromatin structure: histone modification and chromatin remodeling. We asked whether genes involved in these two processes are colocalized in the nucleus. First, we tested whether histone modification regulated genes are colocalized in the nucleus. We collected 25 types of available histone modification data (32,33). Using these data, we identified a cohort of genes that show high modification levels at promoters for each modification (see 'Materials and Methods' section). Using a similar method as above based on hypergeometric test, we found that four modification gene cohorts are colocalized (hypergeometric, $P < 0.01$, after Bonferroni correction; Figure 3A).

Second, we examined whether genes regulated by chromatin remodelers are colocalized in the nucleus. To our knowledge, genome-wide occupancy data are available for eight chromatin remodelers (34). Using these data, we identified cohorts of genes that show high occupancy levels at promoters (near the UAS and TSS, respectively) for each remodeler (see 'Materials and Methods' section). We found that 10 remodeler gene cohorts are colocalized (hypergeometric, $P < 0.01$, after Bonferroni correction; Figure 3B). We next sought to analyze regulatory genes for more chromatin remodelers. As genome-wide occupancy data are available for only a handful of chromatin remodelers, we used a large data set for changes in gene expression accompanying the perturbation (mutation or deletion) of 170 chromatin remodelers (35). If the remodeler regulates chromatin activity of one gene, its perturbation should cause a differential change in expression of this gene because gene expression is linked to chromatin regulation. Genes belong to one remodeler cohort if they display significant changes ($|Z - \text{score}| > 1.64$, $P < 0.05$) in gene expression accompanying the perturbation of the corresponding remodeler. We found that 45 remodeler

gene cohorts are colocalized (hypergeometric, $P < 0.01$, after Bonferroni correction; Figure 3C).

Third, we studied the effect of colocalized chromatin regulation on nucleosomal organization. The nucleosome is the fundamental unit of eukaryotic chromatin. Although nucleosome positioning is partly encoded in genomic sequence (46), it is also determined by chromatin regulators. The observed colocalization of chromatin regulator target genes motivates us to ask whether genes with similar nucleosomal organization are colocalized. A substantial nucleosome-free region (NFR) directly upstream of the TSS is a general feature for most yeast genes (47), but approximately 500 genes exhibit relatively high nucleosome occupancy upstream of the TSS (36). These approximately 500 genes, termed as occupied proximal-nucleosome OPN genes, are colocalized in the nucleus (hypergeometric, $P < 10^{-162}$; Figure 3D). Nucleosome positioning upstream of the TSS on OPN genes are regulated by chromatin regulators, which is consistent with the above observation that genes regulated by chromatin remodelers are colocalized in the nucleus. On the other hand, nucleosomes containing histone variant H2A.Z flank NFR in most promoter regions (48), and their deposition into chromatin requires chromatin remodelers (37). We found that H2A.Z-containing genes are colocalized in the nucleus (hypergeometric, $P < 10^{-67}$; Figure 3D).

Fourth, we examined whether genes with high *trans*-driven gene expression divergence are colocalized. Divergence in gene expression of a specific gene between closely related species can result from sequence changes in its coding region and regulatory region (*cis*), or from changes in sequences or expression of its direct or indirect upstream regulators (*trans*). A recent study has experimentally measured the relative contribution of *cis* and *trans* effects to the gene expression divergence (38). We identified genes with high *trans* and *cis* effects
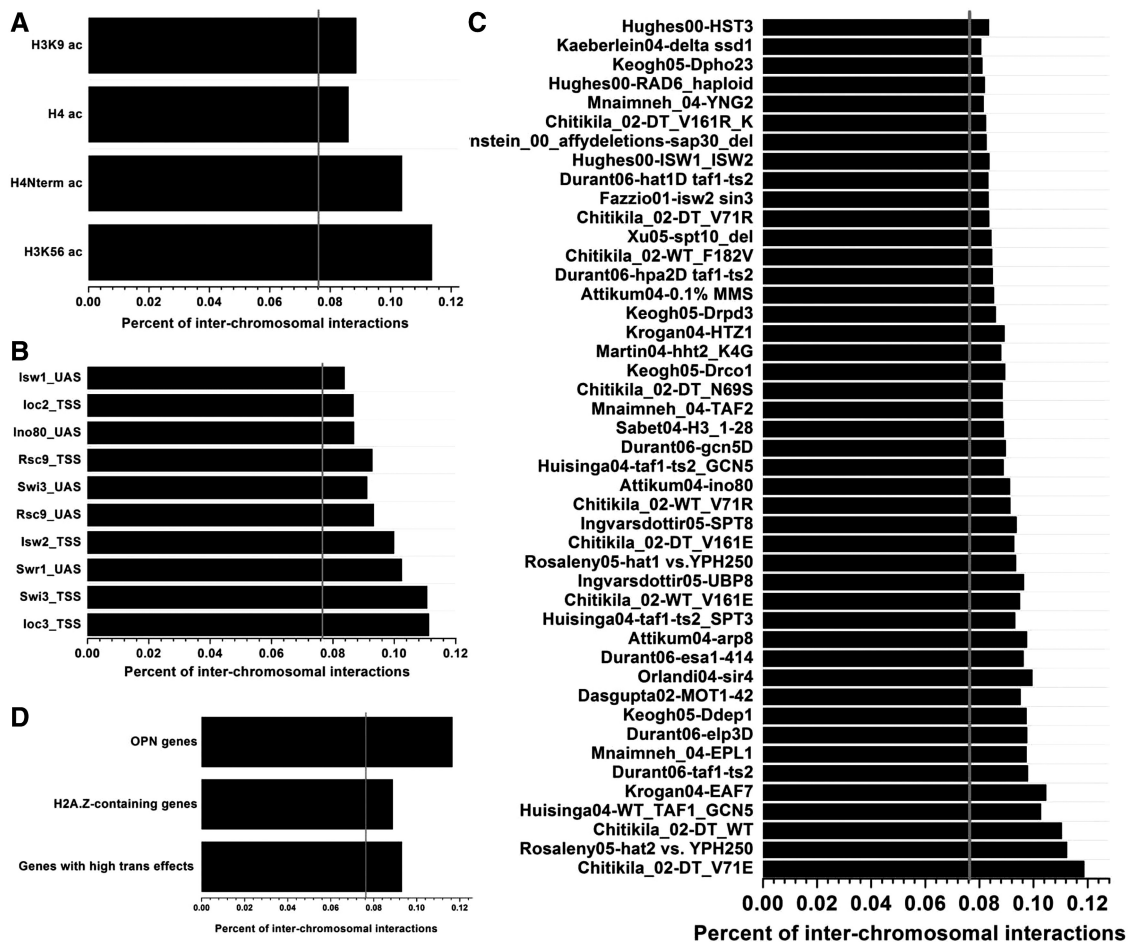
Figure 3. Enrichment of interchromosomal interactions with regard to specific categories of genes. (**A–D**) Each bar represents the enrichment or depletion relative to the percent (gray line) of genome-wide level. Due to lack of space, we only showed genes cohorts that are significantly enriched in interchromosomal interactions (hypergeometric, $P < 0.01$, after Bonferroni correction for multiple testing). (A) Histone modification gene cohorts. Ac represents acetylation. (**B**) Chromatin remodeler occupied gene cohorts. (**C**) Chromatin remodeler knockout affected gene cohorts. The references are listed after the remodeler names. (D) Other gene cohorts.

($|Z - \text{score}| > 1.64$, $P < 0.05$) on the gene expression divergence, respectively. As target genes of *trans* regulators are colocalized, genes with high *trans* effects on gene expression divergence also should be colocalized. Indeed, genes with high *trans* effects are colocalized in the nucleus (hypergeometric, $P < 10^{-34}$; Figure 3D), but genes with high *cis* effects are not (hypergeometric, $P = 0.71$). We also compared the enrichment of interchromosomal interactions for gene cohorts in this section to background levels on individual chromosomes using hypergeometric test (Supplementary Table S2).

Finally, we examined whether the colocalization is linked to particular cellular processes or metabolic or signaling pathways. We tested whether a group of genes are enriched in Gene Ontology (GO) terms or Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using DAVID bioinformatics enrichment tools (49). We found that there is no significant difference in GO terms or KEGG pathways between target-colocalized TFs and the other TFs. For each regulator (i.e. TFs and chromatin remodelers), we statistically examined the enrichment of

its colocalized target genes for each of the GO terms and KEGG pathways by comparing the outcome with that of its non-colocalized target genes (control background), and found that there is no significant difference between them.

DISCUSSION

Here, we investigated genome-wide relationship between gene regulation and gene location in the nucleus. Our major findings are that: (i) a considerable number of TFs show a preference to have targets that are colocalized in the nucleus; (ii) the colocalization of target genes strengthens coregulation of the corresponding TFs; and (iii) most chromatin regulators tend to regulate genes that are colocalized. These findings demonstrate the constraints of gene regulation on genome organization in the nucleus. These results have particular importance for studies attempting to understand gene regulation in the three-dimensional view. Future studies will no doubt need to take gene nuclear location into account in order to achieve a deeper understanding of gene regulation.

Our study provides genome-wide evidence for the colocalization of coregulated genes. A significant number of TFs regulate genes that are colocalized in the nucleus, but target genes of most TFs are not colocalized. It has become evident that the organization of chromatin in the nucleus is not static (50). Examples of dynamic chromatin movements have been documented with regard to gene regulation (51,52). On the other hand, it is well accepted that TFs bind their targets in a dynamic manner. Genes might dynamically move to specific nuclear regions for TF transient binding. Note that the three-dimensional map of yeast genome, we used for analysis represents an 'average' snapshot of genome organization in the nucleus, not the dynamic changes. It is likely that the dynamic colocalization of target genes for most TFs is not captured by this three-dimensional map. It is interesting to test this possibility with the development of experimental methods.

Transcription regulation is a TF-dependent process, and is achieved by the diffusion of TFs. TF regulatory efficiency is influenced by the protein abundance of TFs. If the abundance of one TF is high, it should efficiently regulate its target genes even though targets are scattered over the whole nuclear space. We hypothesized that nuclear colocalization of target-colocalized TF target genes is linked to the low TF abundance. However, we found that the abundance (26) of target-colocalized TFs is comparable with that of the other TFs ($P = 0.67$, Mann–Whitney U-test).

We sought to explain the nuclear colocalization of target genes in terms of the size of target gene cohorts. If one TF regulates a large number of genes, its appropriate regulation could be maintained by the colocalization of its target genes. The nuclear colocalization of target genes should enhance TF regulatory efficiency. However, the sizes of knockout target gene cohorts for target-colocalized TFs are comparable with those of the other TFs (96 versus 110; $P = 0.87$, Mann–Whitney U-test).

TF target genes colocalized are in tighter coregulation by the TFs. TFs identify their target genes by binding the DNA motif sequences in promoter regions. TF binding DNA motifs are usually short and degenerate. There are thus, redundant motifs in the genome, which makes it difficult for TFs to appropriately bind their functional motifs. As gene expression data is measured among a population of cells, the failure of appropriate TF binding in a subpopulation of cells could lead to the apparently weak coregulation of TF target genes. In general, TFs are not randomly distributed in the nucleus, and they show a preference to locate in some distinct nuclear regions. Colocalization of TF target genes in these regions could facilitate the appropriate binding of TFs to function motifs, which strengthens the coregulation of TF target genes.

Histone modification and chromatin remodeling are known to be associated with transcription. It remains to be answered whether the colocalization of their gene cohorts facilitates transcription. We found that histone modifications whose gene cohorts are colocalized in the nucleus, show comparable correlation with transcription activity relative to the other histone modifications ($P = 0.15$, Mann–Whitney U-test; Supplementary Figure S7). Similar result was observed on chromatin remodeler target gene cohorts ($P = 0.49$, Mann–Whitney U-test; Supplementary Figure S7). These results indicate that chromatin regulators whose target genes show colocalization are not significantly associated with transcription activation or repression compared to the other regulators. Moreover, RNA polymerase II-enriched promoters are not colocalized in the nucleus (hypergeometric, $P = 1$). The colocalization of target genes is a feature of some chromatin regulators, but it is not the general feature of chromatin regulators that are linked to transcription activation or repression. The colocalization might facilitate the regulation of chromatin regulators. Its functional roles in biological processes remain to be elucidated.

Our observation that genes coregulated by one chromatin regulator are colocalized has implications. Colocalized genes are regulated by common chromatin regulators. This could result in similar chromatin structures of colocalized genes. Our observation that genes with similar nucleosomal organization are colocalized, supports this concept. In addition, chromatin regulators could move their regulated genes to specialized compartments in the nucleus, leading to the observed colocalization of their coregulated genes. This will give insights into how colocalization of coregulated genes is accomplished.

Despite the findings described above, our study still has limitations. In this study, we used numerous data sets which are measured from different experimental platforms and different yeast strains (Supplementary Table S3). These discrepancies inevitably bias the observations in this study. However, yeast data sets used in this study were all measured in rich media (Supplementary Table S3). The similarity in experimental medium should complement the discrepancies above. It will be of particular interest to perform experiments regarding to this study using the same yeast strains.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

Conflict of interest statement. None declared.

REFERENCES

1. Heun,P., Laroche,T., Shimada,K., Furrer,P. and Gasser,S.M. (2001) Chromosome dynamics in the yeast interphase nucleus. *Science*, **294**, 2181–2186.
2. Casolari,J.M., Brown,C.R., Drubin,D.A., Rando,O.J. and Silver,P.A. (2005) Developmentally induced changes in transcriptional program alter spatial organization across chromosomes. *Genes Dev.*, **19**, 1188–1198.
3. Lanctot,C., Cheutin,T., Cremer,M., Cavalli,G. and Cremer,T. (2007) Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat. Rev. Genet.*, **8**, 104–115.
4. Stone,E.M., Heun,P., Laroche,T., Pillus,L. and Gasser,S.M. (2000) MAP kinase signaling induces nuclear reorganization in budding yeast. *Curr. Biol.*, **10**, 373–382.
5. Gasser,S.M. (2002) Visualizing chromatin dynamics in interphase nuclei. *Science*, **296**, 1412–1416.
6. Cremer,T., Cremer,M., Dietzel,S., Muller,S., Solovei,I. and Fakan,S. (2006) Chromosome territories–a functional nuclear landscape. *Curr. Opin. Cell Biol.*, **18**, 307–316.
7. Cremer,T. and Cremer,C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.*, **2**, 292–301.
8. Reddy,T.E., Pauli,F., Sprouse,R.O., Neff,N.F., Newberry,K.M., Garabedian,M.J. and Myers,R.M. (2009) Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome Res.*, **19**, 2163–2171.
9. Pai,D.A. and Engelke,D.R. (2010) Spatial organization of genes as a component of regulated expression. *Chromosoma*, **119**, 13–25.
10. Lomvardas,S., Barnea,G., Pisapia,D.J., Mendelsohn,M., Kirkland,J. and Axel,R. (2006) Interchromosomal interactions and olfactory receptor choice. *Cell*, **126**, 403–413.
11. Apostolou,E. and Thanos,D. (2008) Virus infection induces NF-kappaB-dependent interchromosomal associations mediating monoallelic IFN-beta gene expression. *Cell*, **134**, 85–96.
12. Spilianakis,C.G., Lalioti,M.D., Town,T., Lee,G.R. and Flavell,R.A. (2005) Interchromosomal associations between alternatively expressed loci. *Nature*, **435**, 637–645.
13. Wansink,D.G., Schul,W., van der Kraan,I., van Steensel,B., van Driel,R. and de Jong,L. (1993) Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. *J. Cell Biol.*, **122**, 283–293.
14. Jackson,D.A., Hassan,A.B., Errington,R.J. and Cook,P.R. (1993) Visualization of focal sites of transcription within human nuclei. *EMBO J.*, **12**, 1059–1065.
15. Osborne,C.S., Chakalova,L., Mitchell,J.A., Horton,A., Wood,A.L., Bolland,D.J., Corcoran,A.E. and Fraser,P. (2007) Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol.*, **5**, e192.
16. Osborne,C.S., Chakalova,L., Brown,K.E., Carter,D., Horton,A., Debrand,E., Goyenechea,B., Mitchell,J.A., Lopes,S., Reik,W. et al. (2004) Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.*, **36**, 1065–1071.
17. Mitchell,J.A. and Fraser,P. (2008) Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev.*, **22**, 20–25.
18. Berger,A.B., Cabal,G.G., Fabre,E., Duong,T., Buc,H., Nehrbass,U., Olivo-Marin,J.C., Gadal,O. and Zimmer,C. (2008) High-resolution statistical mapping reveals gene territories in live yeast. *Nat. Methods*, **5**, 1031–1037.
19. Duan,Z., Andronescu,M., Schutz,K., McIlwain,S., Kim,Y.J., Lee,C., Shendure,J., Fields,S., Blau,C.A. and Noble,W.S. (2010) A three-dimensional model of the yeast genome. *Nature*, **465**, 363–367.
20. Janga,S.C., Collado-Vides,J. and Babu,M.M. (2008) Transcriptional regulation constrains the organization of genes on eukaryotic chromosomes. *Proc. Natl Acad. Sci. USA*, **105**, 15761–15766.
21. Hershberg,R., Yeger-Lotem,E. and Margalit,H. (2005) Chromosomal organization is shaped by the transcription regulatory network. *Trends Genet.*, **21**, 138–142.
22. Lieberman-Aiden,E., van Berkum,N.L., Williams,L., Imakaev,M., Ragoczy,T., Telling,A., Amit,I., Lajoie,B.R., Sabo,P.J., Dorschner,M.O. et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
23. Tanizawa,H., Iwasaki,O., Tanaka,A., Capizzi,J.R., Wickramasinghe,P., Lee,M., Fu,Z. and Noma,K. (2010) Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic Acids Res.*, **38**, 8164–8177.
24. Hu,Z., Killion,P.J. and Iyer,V.R. (2007) Genetic reconstruction of a functional transcriptional regulatory network. *Nat. Genet.*, **39**, 683–687.
25. Harbison,C.T., Gordon,D.B., Lee,T.I., Rinaldi,N.J., Macisaac,K.D., Danford,T.W., Hannett,N.M., Tagne,J.B., Reynolds,D.B., Yoo,J. et al. (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature*, **431**, 99–104.
26. Newman,J.R., Ghaemmaghami,S., Ihmels,J., Breslow,D.K., Noble,M., DeRisi,J.L. and Weissman,J.S. (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature*, **441**, 840–846.
27. Holstege,F.C., Jennings,E.G., Wyrick,J.J., Lee,T.I., Hengartner,C.J., Green,M.R., Golub,T.R., Lander,E.S. and Young,R.A. (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell*, **95**, 717–728.
28. David,L., Huber,W., Granovskaia,M., Toedling,J., Palm,C.J., Bofkin,L., Jones,T., Davis,R.W. and Steinmetz,L.M. (2006) A high-resolution map of transcription in the yeast genome. *Proc. Natl Acad. Sci. USA*, **103**, 5320–5325.
29. Spellman,P.T., Sherlock,G., Zhang,M.Q., Iyer,V.R., Anders,K., Eisen,M.B., Brown,P.O., Botstein,D. and Futcher,B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol. Biol. Cell*, **9**, 3273–3297.
30. Cho,R.J., Campbell,M.J., Winzeler,E.A., Steinmetz,L., Conway,A., Wodicka,L., Wolfsberg,T.G., Gabrielian,A.E., Landsman,D., Lockhart,D.J. et al. (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell*, **2**, 65–73.
31. Gasch,A.P., Spellman,P.T., Kao,C.M., Carmel-Harel,O., Eisen,M.B., Storz,G., Botstein,D. and Brown,P.O. (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell*, **11**, 4241–4257.
32. O'Connor,T.R. and Wyrick,J.J. (2007) ChromatinDB: a database of genome-wide histone modification patterns for Saccharomyces cerevisiae. *Bioinformatics*, **23**, 1828–1830.
33. Pokholok,D.K., Harbison,C.T., Levine,S., Cole,M., Hannett,N.M., Lee,T.I., Bell,G.W., Walker,K., Rolfe,P.A., Herbolsheimer,E. et al. (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell*, **122**, 517–527.
34. Venters,B.J. and Pugh,B.F. (2009) A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. *Genome Res.*, **19**, 360–371.
35. Steinfeld,I., Shamir,R. and Kupiec,M. (2007) A genome-wide analysis in Saccharomyces cerevisiae demonstrates the influence of chromatin modifiers on transcription. *Nat. Genet.*, **39**, 303–309.
36. Tirosh,I. and Barkai,N. (2008) Two strategies for gene regulation by promoter nucleosomes. *Genome Res.*, **18**, 1084–1091.
37. Albert,I., Mavrich,T.N., Tomsho,L.P., Qi,J., Zanton,S.J., Schuster,S.C. and Pugh,B.F. (2007) Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. *Nature*, **446**, 572–576.
38. Tirosh,I., Reikhav,S., Levy,A.A. and Barkai,N. (2009) A yeast hybrid provides insight into the evolution of gene expression regulation. *Science*, **324**, 659–662.
39. Iyer,V.R., Horak,C.E., Scafe,C.S., Botstein,D., Snyder,M. and Brown,P.O. (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature*, **409**, 533–538.

40. Hahn,J.S., Hu,Z., Thiele,D.J. and Iyer,V.R. (2004) Genome-wide analysis of the biology of stress responses through heat shock transcription factor. *Mol. Cell Biol.*, **24**, 5249–5256.

41. Lieb,J.D., Liu,X., Botstein,D. and Brown,P.O. (2001) Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nat. Genet.*, **28**, 327–334.

42. Pilpel,Y., Sudarsanam,P. and Church,G.M. (2001) Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat. Genet.*, **29**, 153–159.

43. Segal,E., Shapira,M., Regev,A., Pe'er,D., Botstein,D., Koller,D. and Friedman,N. (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.*, **34**, 166–176.

44. Cohen,B.A., Mitra,R.D., Hughes,J.D. and Church,G.M. (2000) A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat. Genet.*, **26**, 183–186.

45. Tsai,H.K., Su,C.P., Lu,M.Y., Shih,C.H. and Wang,D. (2007) Co-expression of adjacent genes in yeast cannot be simply attributed to shared regulatory system. *BMC Genomics*, **8**, 352.

46. Segal,E., Fondufe-Mittendorf,Y., Chen,L., Thastrom,A., Field,Y., Moore,I.K., Wang,J.P. and Widom,J. (2006) A genomic code for nucleosome positioning. *Nature*, **442**, 772–778.

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

48. Guillemette,B., Bataille,A.R., Gevry,N., Adam,M., Blanchette,M., Robert,F. and Gaudreau,L. (2005) Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. *PLoS Biol.*, **3**, e384.

49. Huang da,W., Sherman,B.T., Tan,Q., Kir,J., Liu,D., Bryant,D., Guo,Y., Stephens,R., Baseler,M.W., Lane,H.C. *et al.* (2007) DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.*, **35**, W169–W175.

50. Hubner,M.R. and Spector,D.L. (2010) Chromatin dynamics. *Annu. Rev. Biophys.*, **39**, 471–489.

51. Brown,K.E., Baxter,J., Graf,D., Merkenschlager,M. and Fisher,A.G. (1999) Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. *Mol. Cell*, **3**, 207–217.

52. Schubeler,D., Francastel,C., Cimbora,D.M., Reik,A., Martin,D.I. and Groudine,M. (2000) Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus. *Genes Dev.*, **14**, 940–950.