Linking the dynamics of chromatin occupancy and transcription with predictive models

Trung Q. Tran [1], Vinay Tripuraneni [2], Heather K. MacAlpine [2], Sneha Mitra [1], David M. MacAlpine [2,3,⋆], and Alexander J. Hartemink [1,3,⋆]

[1]Department of Computer Science,
Duke University, Durham, NC 27708, USA

[2]Department of Pharmacology and Cancer Biology,
Duke University Medical Center, Durham, NC 27710, USA

[3]Center for Genomic and Computational Biology,
Duke University, Durham, NC 27708, USA

⋆To whom correspondence should be addressed:
Tel: +1 919 681 6077, Email: david.macalpine@duke.edu
Tel: +1 919 660 6514, Email: amink@cs.duke.edu

July 3, 2020

Abstract

Though the sequence of the genome within each eukaryotic cell is essentially fixed, it exists in a complex and changing chromatin state. This state is determined, in part, by the dynamic binding of proteins to the DNA. These proteins—including histones, transcription factors (TFs), and polymerases—interact with one another, the genome, and other molecules to allow the chromatin to adopt one of exceedingly many possible configurations. Understanding how changing chromatin configurations associate with transcription remains a fundamental research problem. We sought to characterize at high spatiotemporal resolution the dynamic interplay between transcription and chromatin in response to cadmium stress. While gene regulatory responses to environmental stress in yeast have been studied, how the chromatin state is modified and how those modifications connect to gene regulation remain unexplored. By combining MNase-seq and RNA-seq data, we found chromatin signatures of transcriptional activation and repression involving both nucleosomal and TF-sized DNA binding factors. Using these signatures, we identified associations between chromatin dynamics and transcriptional regulation, not only for known cadmium response genes, but across the entire genome, including antisense transcripts. Those associations allowed us to develop generalizable models that can predict dynamic transcriptional responses on the basis of dynamic chromatin signatures.

Introduction

Organisms require genic transcription to produce the proteins necessary for biological functions like growth, replication, repair, and response to environmental changes. Transcription is tightly regulated through the complex interplay of a myriad of DNA-binding factors (DBFs), including the histone octamers at the core of a nucleosome, transcription factors (TFs), and polymerases. These proteins and complexes involved in transcription, and the many others interacting with DNA, determine the chromatin landscape. How these constituents of the chromatin bind, unbind, move, and interact to regulate transcription remains an open area of research.

Many studies have made major strides in characterizing the roles of protein complexes involved in transcription. Chromatin immunoprecipitation (ChIP) has been used to assay binding sites of hundreds of proteins on a genomic scale, including factors involved in SAGA-dominated stress-related pathways and TFIID-dominated housekeeping pathways (Venters *et al.* 2011). Likewise, studies have probed proteins involved in the formation of the pre-initiation complex required for transcription initiation (Rhee and Pugh 2012). The role of numerous chromatin remodelers and their interactions have been characterized in detail through ChIP, proteomics, and gene expression analysis of deletion mutants (Krogan *et al.* 2006; Lenstra *et al.* 2011; Shivaswamy and Iyer 2008; Weiner *et al.* 2012, 2015). However, these studies are constrained by limitations in their methods, including lack of antibodies for ChIP and viable deletion strains. Analysis is further complicated by the difficulty in disentangling direct chromatin effects from the pleiotropic action of the many factors and remodelers occurring upstream of transcription. These aspects contribute to why the chromatin landscape involved in transcription is still poorly understood.

An alternative approach has been to profile chromatin occupancy in a protein-

agnostic manner using nuclease digestion. Digestion by a nuclease, such as micrococcal nuclease (MNase), provides a complementary perspective to understand chromatin occupancy as it can probe accessibility at base-pair precision. Recent genome-wide mapping studies have used nucleosome-sized MNase-seq fragments to characterize the dynamics of nucleosomes under various conditions, including the cell cycle (Nocetti and Whitehouse 2016), DNA damage (Tripuraneni *et al.* 2019), and heat shock (Teves and Henikoff 2011). Additionally, studies have attempted to understand the roles of the smaller DNA-bound factors that correspond to subnucleosomal MNase-seq fragments (Belsky *et al.* 2015; Brahma and Henikoff 2019; Chereji *et al.* 2017; Henikoff *et al.* 2011; Kubik *et al.* 2017; Ramachandran *et al.* 2017; Teves and Henikoff 2011). These studies highlight the challenge of characterizing the vast heterogeneity of—and interactions among—proteins and complexes involved in DNA-mediated processes, including transcription.

Factor-agnostic chromatin occupancy profiles from MNase provide an opportunity to link changes in chromatin occupancy at nucleotide resolution with transcriptional regulation, especially regulation induced by environmental perturbations such as cadmium. Cadmium is a toxic metal whose deleterious effects have been well-characterized in yeast. Cadmium toxicity has been shown to induce proteotoxic stress (Faller *et al.* 2005; Gardarin *et al.* 2010; Hartwig 2001; Sharma *et al.* 2008), oxidative stress (Brennan and Schiestl 1996), and inhibition of DNA repair systems (Jin *et al.* 2003). Yeast respond to cadmium exposure through the activation of stress response genes (Dormer *et al.* 2000; Hosiner *et al.* 2014; Jin *et al.* 2008) and the repression of ribosome biogenesis and translation-related genes (Hosiner *et al.* 2014; Jin *et al.* 2008). While these cadmium response pathways have been examined extensively through ChIP, proteomics, and transcription-related studies, the dynamics of a small subset of the chromatin response has only been inferred through deletion mutants or estimated through a limited set of ChIP antibodies.

Utilizing high-resolution spatiotemporal data, we developed general strategies to analyze, genome-wide, chromatin dynamics relative to changes in transcription. We exposed yeast to cadmium and collected data related to its chromatin and gene expression state over a two-hour time course. This data allowed us to infer and differentiate essential cadmium stress response pathways solely from chromatin occupancy alone. We also identified chromatin changes associated with pervasive and potentially regulatory antisense transcription. Our identification of unique classes of chromatin dynamics enabled us to develop a regression model that can predict the yeast's transcriptional response to cadmium.

Results

Paired-end MNase-seq captures high-resolution chromatin occupancy dynamics associated with transcription during cadmium stress

We sought to characterize the dynamics of chromatin in terms of changes in occupancy and organizational structure of nucleosomes as well as smaller transcription-related proteins. A nucleotide resolution view of chromatin occupancy dynamics in response to cadmium stress would allow us to associate and infer relationships between these chromatin changes with changes in transcription. Yeast cells were exposed to cadmium and samples were collected over a two-hour time course (Fig. 1A). Chromatin occupancy and positioning dynamics were profiled using paired-end MNase-seq to map DBFs at base pair resolution (Fig. 1B). Concurrently, transcripts were interrogated using strand-specific total RNA-seq (Fig. 1C).

To evaluate our data and methods, we analyzed the chromatin local to the well-characterized stress response gene HSP26 whose role is to facilitate the disaggregation of misfolded proteins (Cashikar *et al.* 2005). Hsp26 has been implicated in many stress conditions including heat shock (Benesch *et al.* 2010; Franzmann *et al.* 2008), acid-

ity (Kawahata *et al.* 2006), sulfur starvation (Pereira *et al.* 2008), and metals toxicity (Hosiner *et al.* 2014; Momose and Iwahashi 2001). Furthermore, several transcription factors, including Hsf1, Met4, and Met32, have been found to bind in HSP26's well-characterized promoter (Boy-Marcotte *et al.* 1999; Carrillo *et al.* 2012; Chen and Pederson 1993; Susek and Lindquist 1990; Treger *et al.* 1998). The context of these studies makes HSP26 a valuable gene to study in terms of its local chromatin dynamics in an activating condition.

We observe significant changes in the chromatin around HSP26's transcription start site (TSS) (Fig. 2A), coinciding with its increase in transcript level. Upstream, in HSP26's promoter, nucleosome-sized fragments of length 144–174 bp are replaced by small fragments less than 100 bp. In HSP26's gene body, nucleosome-sized fragments become "fuzzy", increasing in positional and fragment-length variability (Fig. 2A). Nucleosomes in HSP26's promoter region are evicted (Lee *et al.* 2004) and replaced by an enrichment of factors associated with transcription initiation pushing nucleosomes downstream (Fig. 2B). Then, active transcription by RNA polymerases displaces and evicts nucleosomes (Kulaeva *et al.* 2010; Lee *et al.* 2004; Schwabish and Struhl 2004), which is apparent in our data in the significant loss of nucleosomal fragments within HSP26's gene body.

Because of these complex transcription-associated chromatin dynamics, we sought to quantify our MNase-seq data by summarizing the chromatin with two scores, which would also enable us to characterize subtle chromatin changes genome-wide, not just at genes with dramatic changes like HSP26. A "promoter occupancy" score was computed by counting the number of small fragments appearing in each gene's promoter region [–200, TSS] for each time point. We used a core promoter length of 200 bp as previously described (Lubliner *et al.* 2013; Smale and Kadonaga 2003). As for the second score, because nucleosomes can be characterized by position changes and variance, in

addition to occupancy, a per nucleotide score was computed using a cross-correlation with a well-positioned, "idealized" nucleosome signal (Fig. 2C). We used the entropy of the +1, +2, and +3 nucleosome cross correlations to define the gene's "nucleosome disorganization". To handle variable RNA stability, we computed transcription rates using difference equations on the transcript levels and previously obtained decay rates (Geisberg *et al.* 2014; Miller *et al.* 2011; Presnyak *et al.* 2015).

Using these measures, we are able to succinctly track the coordination between HSP26's chromatin dynamics and its increased transcription rate (Fig. 2D). Both of HSP26's chromatin scores and its transcription rate increase dramatically throughout the time course and reach their peaks at 120 minutes. In addition to HSP26, we also examined the chromatin at genes with repressed (Supplemental Fig. 1) and unchanging (Supplemental Fig. 2) transcription contexts in response to cadmium and found similar linkages between changes in chromatin and transcription.

We next wanted to determine if this coordination exists genome-wide. The dynamics of the chromatin for each gene were further summarized into a single quantity. For each gene, their scores for promoter occupancy and nucleosome disorganization were transformed into a z-score and combined into an average across the time course (Fig. 3A, B). Sorting by this "combined chromatin" score, we observed a significant proportion of the genome exhibited a coordination between changes in chromatin and changes in transcription for genes that are both activated and repressed. A significant positive Pearson correlation of 0.49 was computed between the average change in promoter occupancy and the average change in transcription (Fig. 3C). Similarly, a correlation of 0.61 was found between the average change in nucleosome disorganization and the average change in transcription (Fig. 3D). When measuring the correlation between the average change in combined chromatin and change in transcription, an even higher correlation of 0.68 was computed (Supplemental Fig. 3A). This correlation

coupled with the lack of correlation (Supplemental Fig. 3B), at 0.17, between promoter occupancy and nucleosome disorganization themselves suggests that each metric provides an orthogonal explanation of the chromatin relative to changes in transcription.

Changes in nucleosome and small factor occupancy at TSSs recapitulate the cell's genome-wide transcriptional response to cadmium

We next sought to determine how well chromatin dynamics reflect the cell's stress response to cadmium exposure. To answer this, we used the three, previously defined scores for chromatin dynamics: promoter occupancy, nucleosome disorganization, and combined chromatin. Using the 300 highest and lowest average scoring genes for each metric, using an approximately 90% inner percentile range, we performed Gene Ontology (GO) enrichment analysis to determine regulation pathways implicated under cadmium exposure. For each chromatin score, discrete GO enrichment pathways were identified with varying levels of false discovery rate (FDR) significance.

A well established response for cells undergoing stress involves shutting down ribosomal and translation-related pathways (Hosiner *et al.* 2014; Reja *et al.* 2015; Vinayachandran *et al.* 2018). We identified this repression through GO enrichment analysis as genes with the greatest decreasing promoter occupancy, nucleosome disorganization, and combined chromatin scores. While, the combined score identifies the translation-related pathways with the greatest significance, many of the terms are recovered with FDR less than 10^{-10} by both promoter occupancy and nucleosome disorganization (Fig. 4A).

While ribosome and translation-related genes are repressed as a tightly regulated cluster, pathways activated under cadmium exposure are recovered with less significance, with FDRs less than 10^{-4} (Fig. 4B). Consistent with cadmium and heavy metals stress studies, two major stress responses are activated under cadmium exposure: sul-

fur assimilation and protein folding (Faller *et al.* 2005; Fauchon *et al.* 2002; Hartwig 2001). Each metric was able to identify distinct pathways with varying FDR. Promoter occupancy implicated sulfur assimilation and response to stress pathways with a FDR of $10^{-3.9}$, and nucleosome disorganization recovered protein refolding with an FDR of $10^{-2.1}$. The combined chromatin score identified GO terms found distinctly in the promoter occupancy or nucleosome disorganization analyses, and for some terms with a better FDR, such as with sulfur amino acid metabolic process.

In both sets of GO analyses, using the combined chromatin score provides novel value in identifying implicated GO terms than using each individual metric alone. And though the cell's response to cadmium has been characterized through gene expression and ChIP-based studies, we show that elements of the chromatin alone are enough to accurately recover major cadmium response pathways.

High-resolution time course recovers cascading induction of sulfur pathways

As suggested by GO enrichment, significant chromatin changes occur at genes in the sulfur metabolic pathways. Utilizing chromatin and transcription data in our time course, we recover findings previously discovered through decades of ChIP, mutant, and transcription-related studies (Barbey *et al.* 2005; Blaiseau and Thomas 1998; Carrillo *et al.* 2012; Cormier *et al.* 2010; Fauchon *et al.* 2002; Kuras *et al.* 1996, 2002; McIsaac *et al.* 2012; Ouni *et al.* 2010; Patton *et al.* 2000; Petti *et al.* 2012) and reveal novel details of the cascade of events regulating these pathways. Yeast cells exposed to cadmium require sulfur metabolized for biosynthesis of the cadmium-chelating glutathione (Fauchon *et al.* 2002). Genes in the sulfur metabolic pathways are activated primarily through the transcription factor Met4 and its binding complex, comprised of cis-binding factors Cbf1 and Met31/Met32, and accessory factor Met28 (Blaiseau and Thomas 1998; Kuras *et al.* 1996). Met4 is regulated through ubiquitination by SCF^{Met30}

either targeting it for degradation or rendering into an inactive, but stable state (Barbey *et al.* 2005; Kaiser *et al.* 2000; Kuras *et al.* 2002). Cadmium exposure overrides this ubiquitination enabling Met4's functional activation of the sulfur metabolic genes (Barbey *et al.* 2005) (Fig. 5A). Using our calculated transcription rates and measures of chromatin dynamics, we recover three major components of the sulfur metabolic pathways (Fig. 5B): (i) activation of the Met4 complex through its cofactors, (ii) activation of the sulfur pathways by Met4, and (iii) implicit down-regulation of Met4 activity by SCF^{Met30}, evident in diminished transcription of Met4-regulated genes.

Upon deubiquitination, Met4 becomes functionally active and induces its own cofactors (Barbey *et al.* 2005; McIsaac *et al.* 2012). Concomitantly, MET32 is activated and, with Met4 through feedforward regulation, activates the sulfur metabolic genes (Carrillo *et al.* 2012; McIsaac *et al.* 2012). In our time course, this induction is evident not only in increased transcription within 7.5 minutes for MET32 and MET28, but also in dramatic nucleosome disorganization in MET32's gene body, (Supplemental Fig. 4) and small fragment enrichment in MET28's promoter. Meanwhile, MET31's chromatin exhibits more unexpected behavior relative to its change in repressed transcription. While Met31 shares a binding motif and largely overlaps in function with Met32 (Blaiseau *et al.* 1997), Met31's role is not as prominent as Met32's in the activation of sulfur pathways (Carrillo *et al.* 2012; McIsaac *et al.* 2012; Petti *et al.* 2012). In our data, we observe that while MET31 is repressed, its nucleosomes disorganize with increased antisense transcription (Supplemental Fig. 5). Additionally, downstream of MET31's transcription end site (TES), small fragments become enriched at a Met31/Met32 binding motif. Taken together, our data identifies MET31 as a potential target for regulation through non-coding RNA (ncRNA) antisense transcription, a result we explore genome-wide in a subsequent section.

Following activation of MET32 and MET28, the Met4 complex is formed and acti-

vates genes in the sulfur pathways (Carrillo *et al.* 2012; McIsaac *et al.* 2012), which we also observe in the form of increased promoter occupancy, nucleosome disorganization, and transcription. Each of the seven sulfur assimilation genes (Fig. 5C) and many of the downstream pathways increase in promoter occupancy and nucleosome disorganization within 15 minutes.

Additionally, Met4 induces a sulfur-sparing transcriptional-switch between functionally similar isoforms to indirectly contribute sulfur required for chelation. This switch includes replacing sulfur-rich Pdc1 with sulfur-lacking Pdc6, Ald4 with Ald6, and Eno1 with Eno2 (Fauchon *et al.* 2002). Each of these sulfur-lacking genes show evidence of activation in both their chromatin and transcription and, consistent with known studies (Fauchon *et al.* 2002), the most dramatic changes are evident as small fragment occupancy changes in PDC6 (Supplemental Fig. 6), PDC1 and ENO1's promoter.

Following the induction of the sulfur pathways, Met32 and Met4's activating functions diminish. Because prolonged activity of Met32 and Met4 induces cell cycle arrest, regulation of Met4 and Met32 through SCF^{Met30} is required for long-term cell proliferation (Ouni *et al.* 2010; Patton *et al.* 2000). These events are present most clearly in the chromatin dynamics for MET30 and MET32. We observe increasing transcription and gradual disorganization of gene body nucleosomes through the 120-minute time course. Additionally, we observe evidence for Met4 facilitating regulation of MET32 (Ouni *et al.* 2010) through a plateau of MET32's nucleosome disorganization and transcription from 60–120 minutes.

Taken together, we are able to detail the timing of the activation of the Met4 complex, induction of the sulfur genes, and subsequent down-regulation of Met4 activity (Fig. 5D). This analysis complements established transcriptional studies by detailing chromatin dynamics of the sulfur metabolic pathways and identifying a potentially

novel regulatory mechanism for MET31 through antisense transcription.

Cadmium treatment induces chromatin dynamics as distinct temporal clusters, including those linked to antisense transcription

Our results demonstrate that the temporal order of chromatin changes were tightly associated with the transcriptional regulation of sulfur pathways. Additionally, we found examples in which chromatin dynamics may not strictly correlate with sense transcription. With these observations in mind, we sought to understand the timing of the chromatin dynamics associated with cadmium stress using hierarchical clustering. We identified eight clusters, to identify generalized patterns, among 832 genes, chosen because they were among the 500 most dynamic in either promoter occupancy or nucleosome disorganization (some genes were in both, which is why there are fewer than 1000; Fig. 6A). GO enrichment analysis was then performed on each cluster. Clustering and GO enrichment analysis revealed three major results: (i) the sulfur and protein folding pathways can be differentiated through the timing of changes in the chromatin, (ii) increased transcription may not always accompany nucleosome disorganization, and (iii) antisense transcription can explain anti-correlated chromatin dynamics.

GO analysis of the eight clusters reveals differences in the timing of the the protein folding and sulfur metabolic pathways (Fig. 6B. Cluster 1, enriched with sulfur assimilation and methionine metabolic process genes, show an increased and steady state of promoter occupancy and nucleosome disorganization, consistent with the previous results. In cluster 2, genes relating to ATPase activity and protein refolding are activated between 7.5–15 minutes, repressed between 30–60 minutes, and reactivated by 120 minutes, evident in the change in promoter occupancy, nucleosome disorganization, and transcription rate for the latter 10% of genes in the cluster, genes that can be differentiated further with a higher k. Clusters 1 and 2 show that not only does the chromatin recover activating stress response pathways, but it also differentiates their

temporal changes consistent with changes in transcription rate.

While most genes presented thus far have shown to have a positive correlation between their changes in each chromatin measure and their change in transcription, clusters 6–8 reveal an unexpected anti-correlated relationship. While the transcription of many of these genes are activated, these genes show either a decrease in promoter occupancy and increase in nucleosome disorganization (in clusters 5 and 6), or an increase in promoter occupancy and decrease in nucleosome disorganization (in clusters 7 and 8). This suggest more complex chromatin dynamics at play than the directly correlated measures we previously described. An example of this complexity is present in cluster 7, where the gene coding for an endoplasmic reticulum membrane protein Mcd4 exhibits chromatin with counter-intuitively organized nucleosomes despite increased transcription. (Supplemental Fig. 7).

For genes in cluster 6, some of the anti-correlated phenomena can be attributed to the antisense transcription (Fig. 6C) previously identified in MET31. Antisense transcription presents itself genome-wide with varying changes in sense transcription (Fig. 7A). We observed two major phenomena consistent with existing studies with respect to antisense transcription. First, as identified in other environmental conditions (Kim *et al.* 2010; Till *et al.* 2018; Wilhelm *et al.* 2008), antisense transcription is induced pervasively in yeast undergoing cadmium stress. The distribution of genes with increased antisense transcripts monotonically skews towards more transcription through the 120-minute time course (Fig. 7B). Of the genes whose sense transcription is unchanging, we found 529 exhibit a four-fold increase in antisense transcription (Fig. 7C). The second phenomena we observed pertains to genes whose sense transcription have changed. Previous studies have found that antisense transcription to be associated with both repression or activation of target genes (Kornienko *et al.* 2013; Swamy *et al.* 2014; Till *et al.* 2018; Vance and Ponting 2014). Under cadmium stress, we identi-

fied 92 genes whose antisense transcripts increased with decreased sense transcription, such as with MET31 and UTR2, whose overexpression has been linked with endoplasmic reticulum stress (Miller *et al.* 2010) (Supplemental Fig. 8). We found 125 genes increased in both sense and antisense transcription, including the gene YBR241C (Supplemental Fig. 9) coding for a vacuole localization protein (Wiederhold *et al.* 2009). These phenomena indicate that changes in the chromatin may not strictly be associated with transcription, at least not solely on the sense strand.

While chromatin dynamics are able to accurately recover and differentiate the timing of known stress response pathways, new questions are raised when these dynamics are not strictly correlated with sense transcription. Complexities introduced by the heterogeneity of DBF binding dynamics and interactions and transcription on the antisense strand indicates that a more complex model of the chromatin landscape is required in elucidating the relationship between chromatin behavior and gene expression.

Chromatin occupancy changes are predictive of changes in gene expression

We next sought to develop a model to quantify the relationship between our measures of the chromatin and changes in transcription. We constructed a Gaussian process regression model to predict the transcription at each time point based solely on chromatin dynamics and the preinduction transcription levels at 0 min. We constructed four models to evaluate various inclusions of measures of the chromatin, including a "full" model incorporating nucleosome positional shift calls (Supplemental Fig. 10) and measures of chromatin relative to called antisense transcripts (Supplemental Fig. 11).

We evaluated each model using 10-fold cross-validation and the coefficient of determination (R^2), as the model's proportion of predictable variance (Fig. 8A). For each model excepting the intercept model, prediction performance becomes worse through the time course as the transcript level deviates from the 0 minute transcript

feature. However, models including features of the chromatin consistently outperform the model using 0 min transcript level alone. Nucleosome disorganization is more informative than promoter occupancy and, consistent with previous results, combining both metrics provides more predictive power than each alone. The full model is not the best between 7.5–15 minutes because prediction is mainly driven by 0 minute transcript level early on (Fig. 8B). It outperforms all other model between 30–120 minutes maintaining an R^2 greater than 0.4 two hours after the cell's exposure to cadmium (Fig. 8C, D).

While our metrics do not describe the full state and variability of the chromatin landscape during transcription, our regression model provides evidence that a proportion of transcription can be explained from modifications of the chromatin state. This model serves as a baseline for understanding a portion of the complex relationship between the chromatin and transcription with numerous opportunities for extension.

Discussion

In contrast to ChIP-based studies, our study surveys the occupancy of DBFs across the entire genome without explicit information on the identities of the DBFs. While nucleosomes are well-characterized by nuclease digestion studies, profiling TFs and complexes that affect gene expression is a more challenging, open problem. Studies have identified the dynamics of various promoter-binding factors including transcription factors, general transcription factors, polymerases, mediator, SAGA, TFIID, histone modifications, chromatin remodelers, and others (Chereji *et al.* 2017; Huisinga and Pugh 2004; Reja *et al.* 2015; Rhee and Pugh 2012; Shivaswamy and Iyer 2008; Venters *et al.* 2011; Vinayachandran *et al.* 2018; Weiner *et al.* 2012, 2015). Utilizing both literature and motif analysis of TFs, we can implicitly describe the activity evident within gene promoters, such as with HSP26. Additionally, well-characterized responses, such as the

sulfur pathways, allow for additional context in determining the logical sequencing of chromatin modification events and modal changes in gene expression.

Analysis of the gene encoding Met4 cofactor Met31 uncovered chromatin changes, not only linked with gene expression, but also with antisense transcription. While pervasive and regulatory ncRNA and antisense transcription have previously been shown to be associated with environmental perturbation (Camblong *et al.* 2007; Nadal-Ribelles *et al.* 2014; Swamy *et al.* 2014; Toesca *et al.* 2011), we characterized the relationship between these transcripts with gene expression from the perspective of the chromatin. Including the chromatin measures for the 667 genes with antisense transcripts also provides a marginal benefit in predicting sense transcripts (Fig. 8A). This benefit can be explored further by narrowing in on the effect size of these antisense-related chromatin measures and by examining the individual sets of genes whose gene expression appears to have a relationship with antisense transcription.

Using the initial transcript level and chromatin dynamics of both sense and antisense transcription, our regression model is able to predict the level of sense transcript with an R^2 greater than or equal to 0.44 for all time points following cadmium exposure Fig. 8A. There are multiple opportunities to extend this model. Further quantifying of the chromatin may include additional classes of fragments and characterization of the chromatin outside our defined [–200, 0] bp promoter and [0, 500] bp gene body. Additionally, this data set enables opportunities for modeling using other statistical methods including generalized linear models, deep neural networks, or random forests. This model and its predictions serve as a baseline showing the potential modeling opportunities and richness of statistical power of the chromatin.

Materials and Methods

Yeast strain

The yeast strain used in this study has the W303 background with the genotype: MATa, leu2-3,112, trp1-1, can1-100, ura3-1, ade2-1, his3-11,15.

Cell growth

Cells were grown asynchronously in rich medium at 30°C to an OD_{600} of 0.8. A sample was removed and crosslinked for MNase-seq and another was pelleted and flash frozen for RNA-seq at time 0 before the addition of $CdCl_2$ to a final concentration of 1mM. Samples were taken at 7.5 min, 15 min, 30 min, 60 min, and 120 min following $CdCl_2$ addition. All samples were taken and processed in duplicate.

Chromatin preparation

Chromatin was prepared as previously described (Belsky *et al.* 2015).

RNA-seq

RNA was prepared using the Illumina TruSeq Stranded Total RNA Human/Mouse/Rat kit (Cat number RS-122-2201) following the protocol provided by Illumina with Ribo-Zero.

Sequencing library preparation

Illumina sequencing libraries of MNase-treated DNA were prepared using 500 ng of DNA as previously described (Henikoff *et al.* 2011).

Sequencing read alignment to the genome

All reads were aligned to the sacCer3/R64 version of the *S. cerevisiae* genome using Bowtie 0.12.7 (Langmead *et al.* 2009). The recovered sequences from all paired-end MNase reads were truncated to 20bp and aligned in paired-end mode using the following Bowtie parameters: `-n 2 -l 20 -m l -k l -X 1000`.

MNase-seq duplicates, A and B, were randomly subsampled and merged to reduce bias from library preparation, sequencing, and MNase digestion.

$$A = \{a_0, a_{7.5}, a_{15}, a_{30}, a_{60}, a_{120}\}$$

$$B = \{b_0, b_{7.5}, b_{15}, b_{30}, b_{60}, b_{120}\}$$

Separately for each duplicate, the time point with the fewest reads determined the subsampling depth (k_A and k_B):

$$T = \{0, 7.5, 15, 30, 60, 120\}$$

$$k_A = \min_{t \in T} |a_t| \ , \ k_B = \min_{t \in T} |b_t|$$

Each duplicate was then subsampled (uniformly at random) to its respective subsampling depth to form new sets A' and B':

$$A' = \{a'_0, \ldots, a'_{120}\}$$

$$B' = \{b'_0, \ldots, b'_{120}\}$$

$$|a'_t| = k_A \ , \ |b'_t| = k_B \ , \ \forall t \in T$$

Finally, the subsampled duplicates were merged into a superset, M, for downstream analysis:

$$M = \{m'_0, \ldots, m'_{120}\} = \{a'_0 \cup b'_0, \ldots, a'_{120} \cup b'_{120}\}$$

Selection of gene set

We compiled a set of 4,427 genes for analysis. Genes were chosen according to five criteria: (i) classified as either verified or uncharacterized by sacCer3/R64, (ii) contains an open reading frame (ORF) at least 500 bp long, (iii) contains an annotated TSS, (iv) has an estimated half-life value, and (v) has adequate MNase-seq coverage.

Genes whose ORFs are less than 500 bp (Supplemental Fig. 12A) long were omitted in order to ensure valid "gene body" calculations between [TSS, +500]. TSS annotations were determined by (Park *et al.* 2014). For four important sulfur-related genes, Sul1, Sul2, Met32, and Hsp26, TSS annotations were manually annotated to be consistent with this study's RNA-seq data. A half-life was required for each gene in order to estimate valid transcription rates. MNase-seq coverage was computed in a 2,000 bp window centered on each gene's TSS. A position in this window is considered "covered" when there exists at least one read whose center lies on this position. MNase coverage was then defined as the number of covered positions in this window divided by the the length of the window, 2,000 bp. Genes with MNase coverage below 0.85 (n=109) were excluded from further analysis (Supplemental Fig. 12B).

Classification of MNase-seq fragments and occupancy

For each gene, two regions were defined relative to their TSS. The promoter region was defined as a 200 bp region upstream of the TSS, [–200, TSS]. This region was chosen as a length previously described by (Lubliner *et al.* 2013; Smale and Kadonaga 2003). The gene body region was defined as the 500 bp region downstream of the TSS, [TSS, +500], to include the +1, +2, and +3 nucleosomes.

To compute metrics against nucleosome and small factor binding signals, two reference data sets were used. Nucleosome-related metrics were computed by examining

the MNase-seq fragment distribution at 2,500 unique nucleosome positions mapped by a highly sensitive chemical mapping methodology (Brogaard *et al.* 2012). Small factors metrics were computed using 279 Abf1 binding sites determined through phylogenetic conservation and motif discovery obtained from http://fraenkel-nsf.csbi.mit.edu/improved_map/p00 (MacIsaac *et al.* 2006). Prior studies have found clear signals of small MNase-seq fragment enrichment at Abf1 sites (Henikoff *et al.* 2011). MNase-seq distributions at each reference set was examined at 0 minutes, prior to cadmium treatment.

Reads were further delineated as nucleosome-sized fragments, those between 144–174 bp long, or small fragments, those less than 100 bp long. A mode length of 159 bp was computed at the MNase-seq fragments at (Brogaard *et al.* 2012) sites. A ± 15 bp range around this 159 bp mode length was chosen for nucleosome-sized fragments. A mode length of 75 bp was computed at (MacIsaac *et al.* 2006) sites. Fragments less than 100 were chosen to be small fragments. Occupancy was then defined by counting the number of fragment centers in these regions with the designated fragment lengths. Occupancy was calculated for small fragments in the promoter, nucleosome fragments in the promoter, small fragments in the gene body, and nucleosome fragments in the gene body.

Signal processing of chromatin

A cross-correlation was computed in a similar manner described in (Tripuraneni *et al.* 2019). Around each gene's TSS, a per bp cross-correlation score was computed to smooth the positional variation and filter out non-relevant fragments. Three two-dimensional cross-correlation kernels were constructed, an "idealized", well-positioned nucleosome (Supplemental Fig. 13A) kernel, a clearly bound small factor kernel (Supplemental Fig. 13B), and a triple nucleosome "gene body" summarization kernel (Supplemental Fig. 13C). Each kernel was applied to the region local to each gene's TSS for

each time point to compute a per bp cross-correlation score (Supplemental Fig. 13D).

The nucleosome and small factor kernels were constructed using a bivariate Gaussian distribution parameterized by the mean and variance for the position and length for MNase-seq fragments. The parameters for each kernel were determined using the fragment length and position distributions at positions in (Brogaard *et al.* 2012) and (MacIsaac *et al.* 2006) previously described in *Classification of MNase-seq fragments and occupancy*.

To summarize the gene body chromatin as a whole, a three nucleosome, "triple" kernel was constructed to dampen the effect of the +1 nucleosome becoming more poised to be well-positioned (Nocetti and Whitehouse 2016). The triple nucleosome kernel was constructed by repeating the nucleosome kernel and increasing the variance to take into account variable linker spacing. The nucleosome kernel spacing was determined using the average peak spacing between the [+1,+2] and the [+2,+3] nucleosome cross correlation scores Supplemental Fig. 13E.

Quantifying nucleosome disorganization

For each gene, a random variable X was defined with n possible outcomes representing each position to evaluate relative to the gene TSS.

$$X = \{1, \ldots, n\}$$

The probability of each outcome is estimated using the triple nucleosome cross correlation scores previously defined and normalized to sum to 1.

Because the triple kernel computes a score for three approximately adjacent nucleosome positions, we set $n = 150$ to summarize the disorganization of the first three

nucleosomes in the gene body starting with +1 within the [0, 150] window.

$$\text{cross}_{\text{nuc}}(i) = \text{nucleosome cross correlation at } i$$
$$\hat{p}(X = x_i) = \frac{\text{cross}_{\text{nuc}}(i)}{\lambda}, \qquad \text{where } \lambda = \sum_{i}^{n} \text{cross}_{\text{nuc}}(i)$$

Using this random variable, a score was computed for each gene to define its "nucleosome disorganization" using information entropy (Supplemental Fig. 13F):

$$H(X) = -\sum_{i \in 1 \ldots n} P_X(x_i) \cdot \log_2 P_X(x_i)$$

Calling +1, +2, +3 nucleosomes

Nucleosomes were called using the peak nucleosome cross correlation scores local to each gene's TSS. The peak scores per bp of a 1000 bp window around the TSS were sorted, and with the largest peak iteratively removed. Positions within 150 bp around each peak were also removed and this value and position was called as a nucleosome. This procedure was repeated until all positions were removed and nucleosomes were called for this 1000 bp window.

"Linked" nucleosomes are defined as nucleosomes across the time course that nominally represent the same underlying nucleosome that may have changed in position or "fuzziness". Nucleosomes were linked across time points using a nearest neighbor approach. In a greedy manner, the nucleosome the lowest disorganization score, the most well-positioned, was considered first. The position of this nucleosome was used to identify the linked nucleosomes in previous and subsequent time points by considering the nearest nucleosome for their respective time points within 100 bp of the original nucleosome's position.

+1 nucleosomes were called by identifying linked nucleosomes closest to the TSS. +2 and +3 nucleosomes were computed as the next set of nucleosomes at least 80 bp downstream from their neighboring nucleosome.

Gene Ontology enrichment analysis

GO enrichment analysis was performed using GOATOOLS (Klopfenstein *et al.* 2018) with the go-basic.obo annotations from the Gene Ontology Consortium (Ashburner *et al.* 2000; The Gene Ontology Consortium 2019). False discovery rate was correcting using the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995).

Identification of transcription factor binding sites

TF binding sites were identified with FIMO (Grant *et al.* 2011) using the motif database from MacIsaac *et al.* (2006) and default p-value threshold. Selected binding sites with supporting literature were annotated on typhoon plots.

Transcription rate estimation

As previously described in (Cashikar *et al.* 2005; Rabani *et al.* 2011; Yang *et al.* 2003), transcription rates were computed for each gene using a zero-order growth with first-order decay relationship:

$$\begin{aligned} \frac{dC_i}{dt} &= R_i - k \cdot C_i \\ C_i &= \frac{R_i}{k} + G_i \cdot e^{-kt_i} \\ t_i &\in \{7.5, 15, 30, 60, 120\}, \text{ s.t. } i \in \{1, \ldots, 5\} \end{aligned}$$

Where C_i is the total RNA concentration measured by RNA-seq for sample i, k is the fixed decay rate, G_i is the concentration of RNA affected by the zero order growth, and R_i is the unknown transcription rate. Assuming a constant rate of transcription between time points, R_i and G_i can be solved through pairs of difference equations:

$$\begin{aligned} C_{i-1} &= \frac{R_i}{k} + G_i \cdot e^{-k \cdot t_{i-1}} \\ C_i &= \frac{R_i}{k} + G_i \cdot e^{-k \cdot t_i} \end{aligned}$$

Transcription rates for $R_i, i \in \{7.5, 15, 30, 60, 120\}$ were computed as systems of difference equations between pairs of RNA-seq measurements to compute G_i and R_i.

Similarly, steady-state transcription rates, R_0 at t_0 were computed by setting the rate of production equal to the rate of decay:

$$t_0 = 0, i = 0$$

$$R_0 = k \cdot C_0$$

Decay rates were computed using the average half-life values, τ between (Geisberg *et al.* 2014; Miller *et al.* 2011; Presnyak *et al.* 2015) and $k = 1/\tau$. Computed transcription rate values were then truncated to 0.1 TPM/min for valid fold-change evaluation.

Clustering of chromatin measures

Clustering was performed using hierarchical clustering through SciPy (Virtanen *et al.* 2020) for its flexibility in determining k. The Ward linkage was used for its efficient approximation to the minimal sum of squares objective (Ward 1963).

832 genes were chosen for clustering from the union of the 500 greatest increase in average promoter occupancy or 500 greatest increase in average nucleosome disorganization, genes outside of an approximately 75% inner percentile range for each measure.

Clustering was performed against the pair-wise Euclidean distances between the z-score normalized measures of change in promoter occupancy and nucleosome disorganization. Clustering to k=8 was chosen to balance the interpretability of fewer clusters with the significance of identified GO terms in smaller, but more numerous clusters.

Computing antisense transcription metrics

Antisense transcript levels were quantified using a TPM calculation defined by (Wagner *et al.* 2012) for strand-specific RNA-seq reads on the antisense strand in the ORF for each gene.

TSSs and transcription end sites (TESs) for antisense transcripts were identified using RNA-seq pileup, the number of reads covering a genomic position. To increase the signal of fully transcribed transcripts, per-position pileup values were summed across each time point into a cumulative pileup and smoothed using a Gaussian kernel.

Antisense transcripts were identified starting with the highest cumulative pileup value within a gene's ORF on the antisense strand. The antisense TSS and TES were each identified by progressively searching upstream and downstream to identify the positions in which the cumulative pileup values are minimized (Supplemental Fig. 11A). Antisense transcripts were not called if they did not meet a minimum threshold.

For the 667 genes in which an antisense transcript could be called (Supplemental Fig. 11B), nucleosome disorganization and promoter occupancy measures were computed, as previously described on the sense strand, relative to the antisense TSSs.

Transcript level prediction model

Gaussian process regression models were constructed to predict the $\log_2$ transcript level for each time point using the $\log_2$ transcript level at time 0, features of the chromatin at 0 minutes, and features of the chromatin for the time being predicted.

Four models were constructed to compare various combinations of measures of the chromatin: a small fragments promoter occupancy model, a gene body nucleosome disorganization model, a combined chromatin model, and a full model incorporating

all previous models' features with the addition of +1, +2, and +3 nucleosome position shift relative to 0 min (Supplemental Fig. 10) and measures of chromatin relative to called antisense transcripts (Supplemental Fig. 11).

Each Gaussian process regression model developed using `scikit-learn` (Pedregosa *et al.* 2011) with a radial-basis function (RBF) kernel with length scale bounded between 0.1 and 100 and a white kernel with noise level 10^{-4} as priors for covariance. The length scale bounds and noise parameters were determined empirically through a sensitivity analysis on a subset of the data.

Promoter occupancy and nucleosome disorganization measures were log transformed to an approximately normal distribution. Then, each chromatin measure, including nucleosome shift, was z-score normalized so that the RBF length parameter could be successfully approximated.

Performance for each model was evaluated using the coefficient of determination, R^2, under 10-fold cross validation.

Data Accession

Code for this work is available at https://github.com/HarteminkLab/cadmium-paper. Data is available on NCBI/GEO with accession number GSE153609.

Acknowledgments

The authors would like to thank Yulong Li and Greg Crawford for their suggestions and helpful comments during the development of this research. The work was supported by the following grants from the NIH National Institute of General Medical Sciences:

R35-GM127062 (D.M.M.) and R01-GM118551 (A.J.H.).

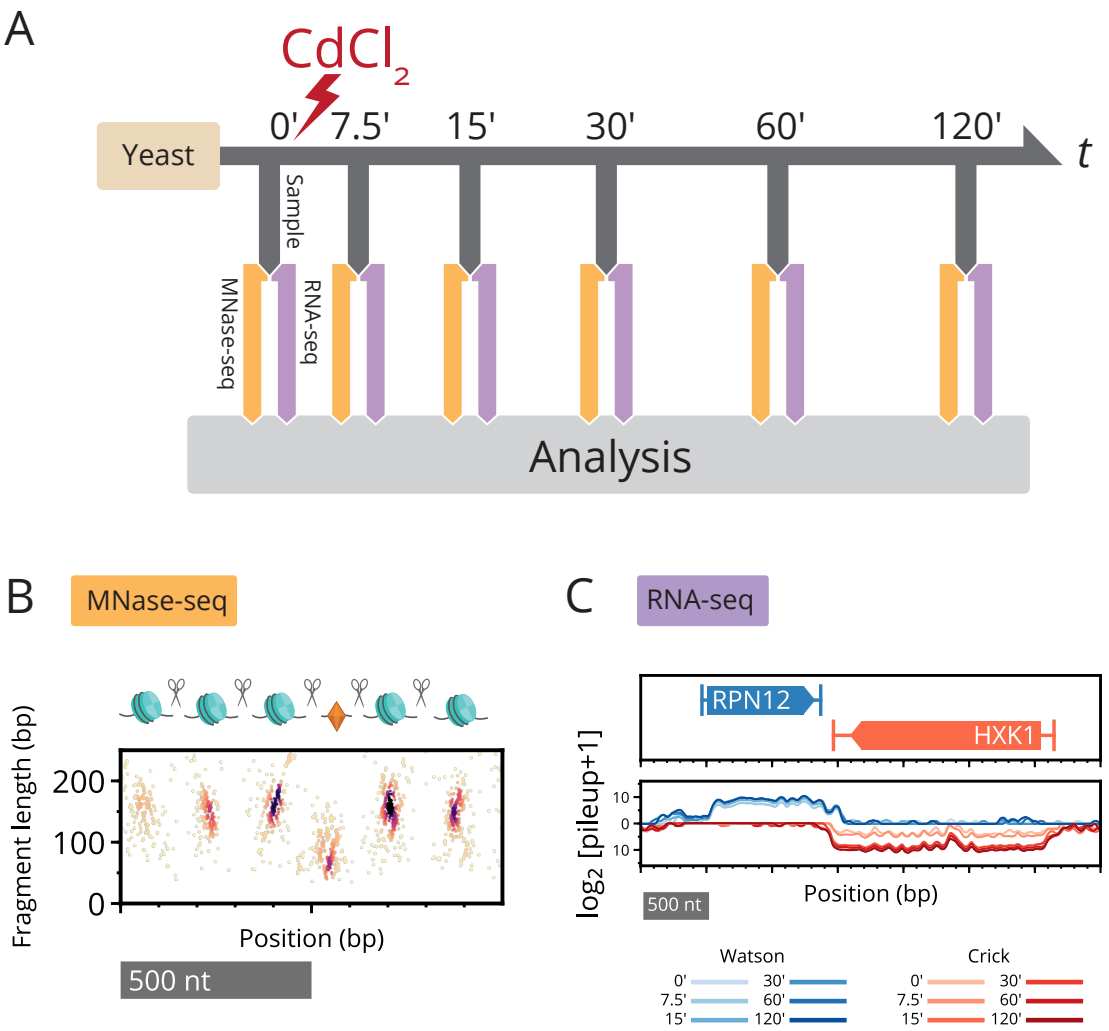

Figure 1. Paired-end MNase-seq and stranded RNA-seq capture high-resolution chromatin occupancy and transcriptome state throughout a perturbation time course. **(A)** Overview of cadmium perturbation experiment in which paired-end MNase-seq and strand-specific RNA-seq samples were collected immediately prior to cadmium exposure and for five additional time points over two hours. **(B)** Depiction of nucleosomes flanking a small (subnucleosomal) binding factor, and fragments that result upon digestion by MNase. Paired-end MNase-seq fragments are plotted based on their center position and length. **(C)** Strand-specific RNA-seq is plotted as the $\log_2$ pileup, the number of total RNA-seq reads at each genomic position, separately mapped to Watson (blue) and Crick (red) strands. Changing RNA-seq read levels over the time course are plotted using progressive coloring for each strand.

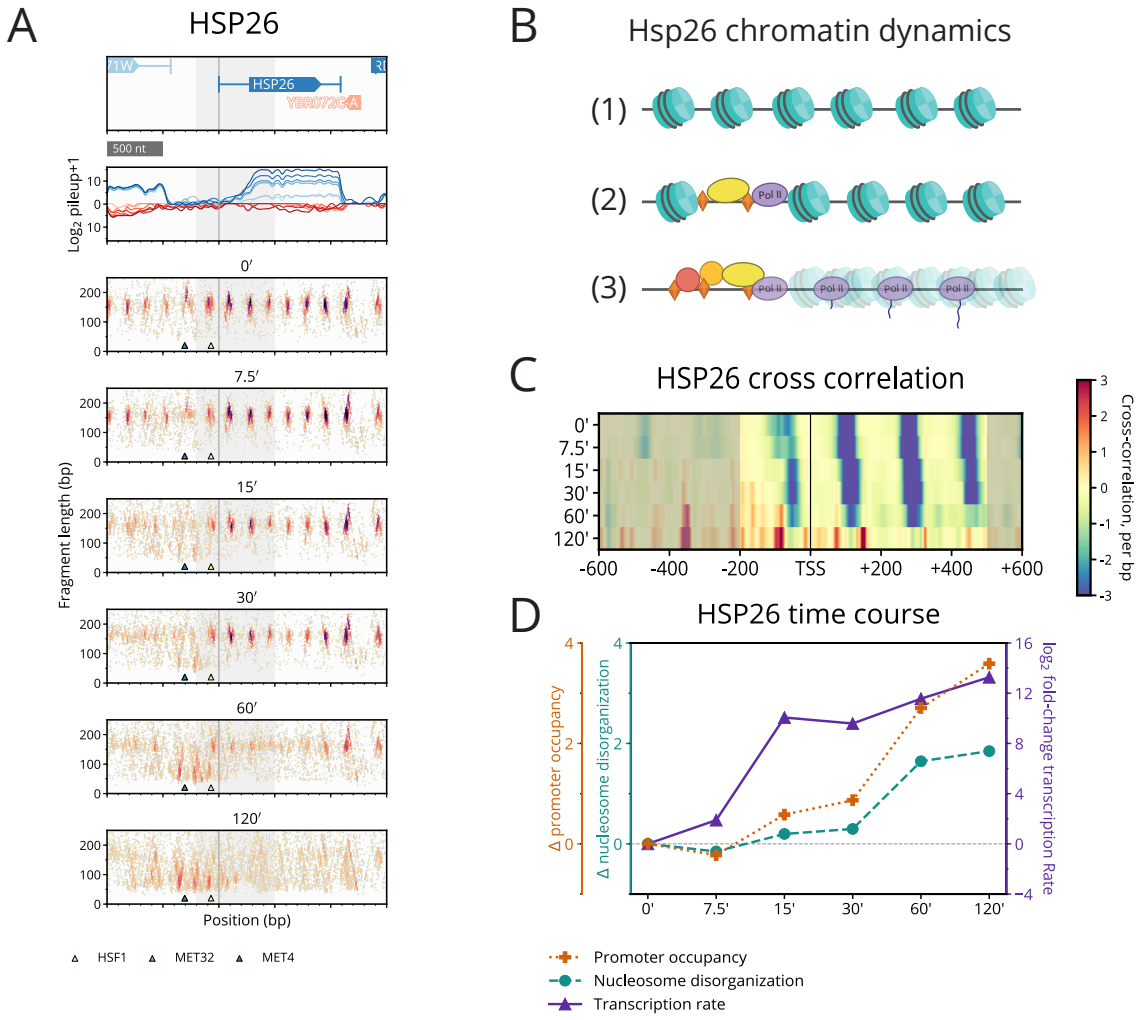
A
HSP26
HSP26
YBR072C-A
500 nt
Log2 pileup+1
0′
7.5′
15′
30′
60′
120′
Fragment length (bp)
Position (bp)
HSF1
MET32
MET4
B
Hsp26 chromatin dynamics
(1)
(2)
(3)
Pol II
C
HSP26 cross correlation
0'
7.5'
15'
30'
60'
120'
-600
-400
-200
TSS
+200
+400
+600
Cross-correlation, per bp
D
HSP26 time course
Δ promoter occupancy
Δ nucleosome disorganization
log2 fold-change transcription Rate
Promoter occupancy
Nucleosome disorganization
Transcription rate

Figure 2. Cadmium induces local chromatin dynamics that correlate with transcription of HSP26. **(A)** Typhoon plot shows dynamics of MNase-seq and RNA-seq data near HSP26. Nucleosomes in the promoter region are replaced by small fragments, while gene body nucleosomes disorganize (grey shading highlights the [–200,500] region around the TSS that we analyze for all genes). Small fragments appear around motifs for known regulators Hsf1 (yellow triangle), Met4 (blue triangle), and Met32 (obscured by blue triangle) **(B)** Depiction of HSP26's chromatin dynamics. (1) Before treatment, nucleosomes are well-positioned. (2) Between 15–30 minutes, nucleosomes are evicted from the promoter region and replaced by transcription-related proteins and complexes. (3) By 60–120 minutes, nucleosomes are fuzzy and polymerases are actively transcribing HSP26. **(C)** Heatmap of differential cross-correlation values of HSP26 through time course, summarizing how gene body nucleosomes initially shift downstream and then disappear, and how promoter nucleosomes are rapidly displaced as small fragments accumulate. Higher values (more red) indicate higher cross-correlation with subnucleosome fragments; lower values (more blue) indicate a stronger signal for nucleosome fragments. **(D)** Line plot of HSP26 time course summarizing the change relative to 0 min in occupancy of promoter small fragments (orange), disorganization of gene body nucleosomes (turquoise), and transcription rate (purple).

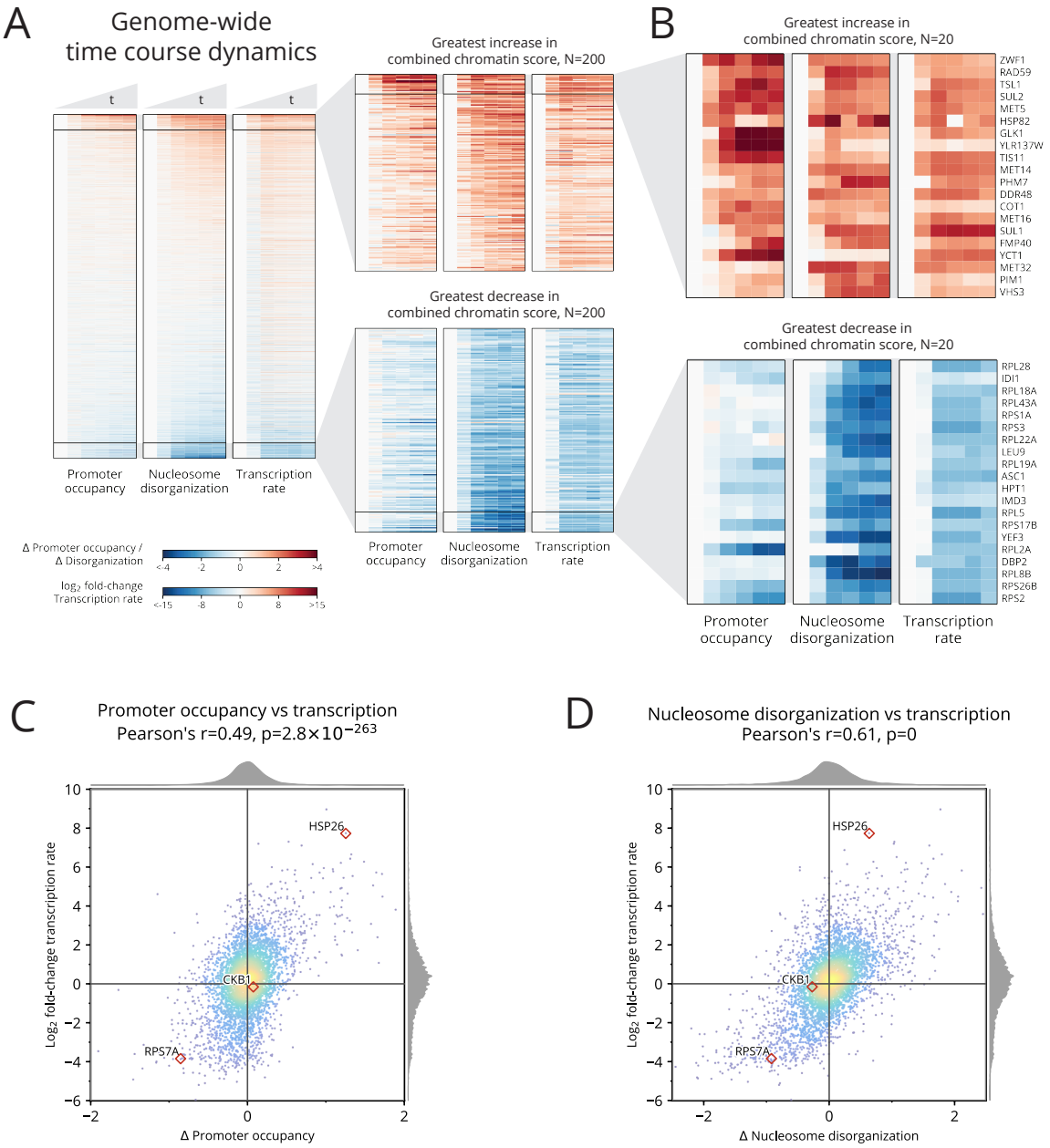

Figure 3. Cadmium induces genome-wide chromatin dynamics that correlate well with genome-wide transcriptional dynamics. **(A)** Heatmaps of changes in chromatin occupancy measures and transcription rate for all genes and all times, relative to 0 min (left: promoter small fragment occupancy; middle: gene body nucleosome disorganization; right: transcription rate). Genes (rows) are sorted by combined chromatin score. **(B)** Detailed heatmaps of the 20 genes whose combined chromatin scores increase (top) or decrease (bottom) most. **(C)** Scatter plot of relationship between change in promoter occupancy and change in transcription rate, each averaged over the time course. **(D)** Scatter plot of relationship between change in nucleosome disorganization and change in transcription rate, each averaged over the time course.

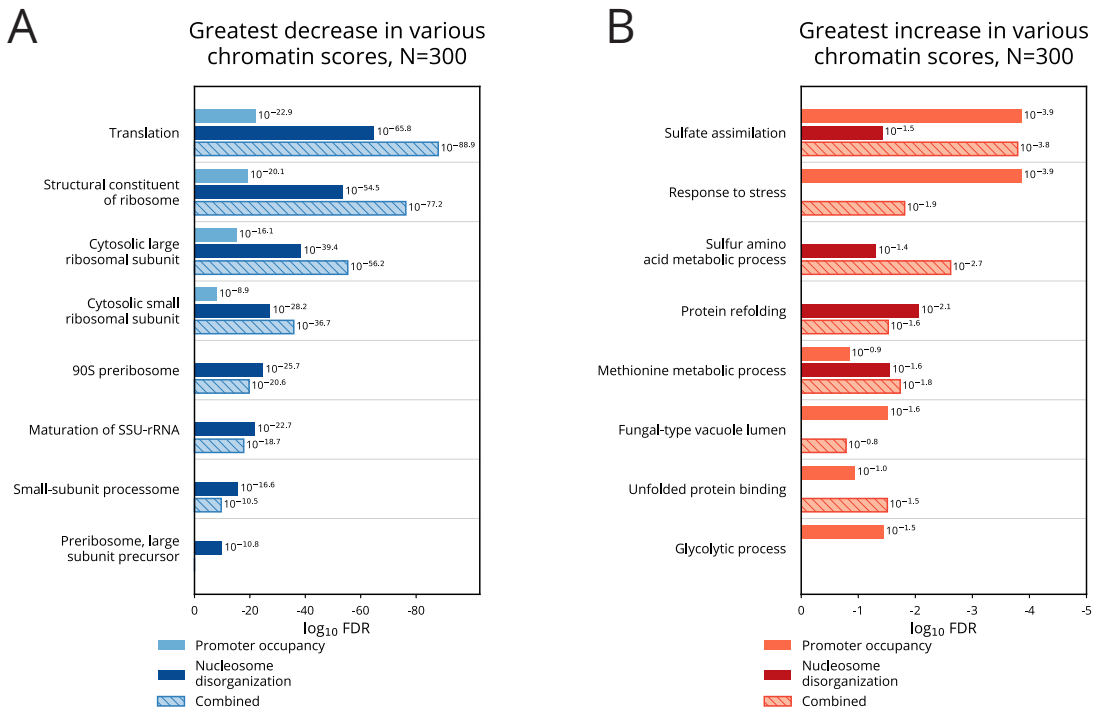

Figure 4. GO enrichment analysis of genes with highly dynamic chromatin support established cadmium response pathways. **(A)** GO enrichment analysis of 300 genes with greatest decrease in promoter occupancy, nucleosome disorganization, and combined chromatin score. Translation-related genes are recovered with significant FDR. **(B)** GO enrichment analysis of 300 genes with greatest increase in promoter occupancy, nucleosome disorganization, and combined chromatin score. Genes involved with stress response, sulfur assimilation, and protein folding pathways are recovered with significant FDR.

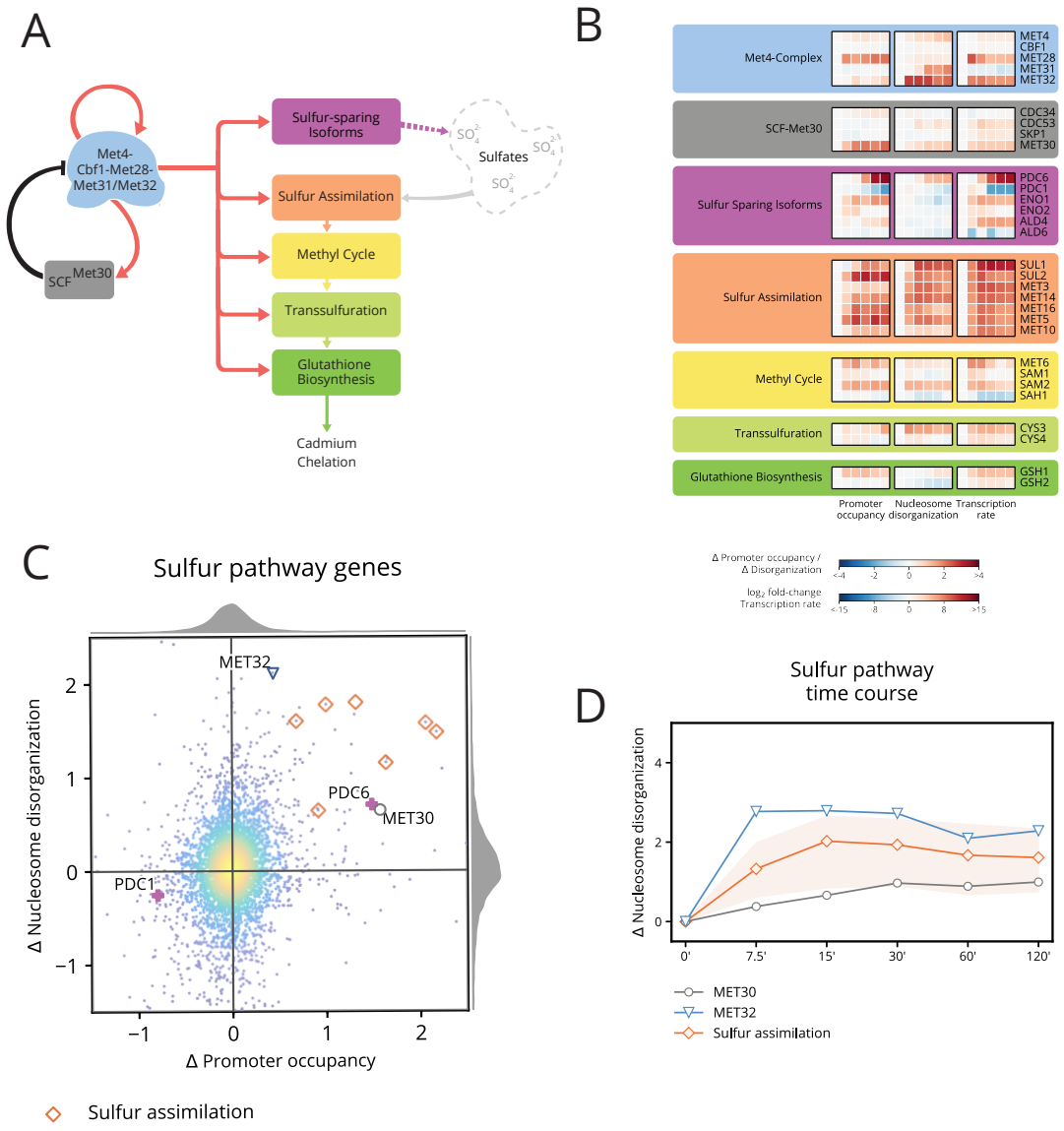
A
Met4-Cbf1-Met28-Met31/Met32
SCF Met30
Sulfur-sparing Isoforms
Sulfur Assimilation
Methyl Cycle
Transsulfuration
Glutathione Biosynthesis
Sulfates
Cadmium Chelation
B
Met4-Complex
MET4 CBF1 MET28 MET31 MET32
SCF-Met30
CDC34 CDC53 SKP1 MET30
Sulfur Sparing Isoforms
PDC6 PDC1 ENO1 ENO2 ALD4 ALD6
Sulfur Assimilation
SUL1 SUL2 MET3 MET14 MET16 MET5 MET10
Methyl Cycle
MET6 SAM1 SAM2 SAH1
Transsulfuration
CYS3 CYS4
Glutathione Biosynthesis
GSH1 GSH2
Promoter occupancy
Nucleosome disorganization
Transcription rate
Δ Promoter occupancy / Δ Disorganization
log2 fold-change Transcription rate
C
Sulfur pathway genes
MET32
PDC6
MET30
PDC1
Δ Nucleosome disorganization
Δ Promoter occupancy
Sulfur assimilation
D
Sulfur pathway time course
Δ Nucleosome disorganization
0'
7.5'
15'
30'
60'
120'
MET30
MET32
Sulfur assimilation

Figure 5. Chromatin and transcription dynamics detail Met4 and Met32 functional activation, induction of sulfur genes, and subsequent regulation. **(A)** Met4 complex activates cascading sulfur pathways required for cadmium chelation and is regulated by SCF^{Met30} **(B)** Heatmap of changes in chromatin occupancy and transcription rate for the sulfur pathway genes. Cofactors of the Met4 complex exhibit dramatic chromatin changes in promoter occupancy (for MET28) and nucleosome disorganization (for MET32). Sulfur sparing isoforms exhibit inverse chromatin dynamics most pronounced between PDC6 and PDC1. Nearly all of the sulfur assimilation pathway have a dramatic increase in promoter small fragments and nucleosome disorganization. **(C)** Scatter plot of the average change in promoter occupancy and average change in nucleosome disorganization. Chromatin dynamics in sulfur-related genes may appear primarily through a change in a single measure of the chromatin, as with MET32 (blue triangle), MET30 (gray circle), and PDC6/PDC1 (violet), or as changes in both promoter occupancy and nucleosome disorganization, such as with the sulfur assimilation genes (orange diamonds) **(D)** Line plot of the change in nucleosome disorganization for the regulator gene MET30, activator MET32, and the sulfur assimilation genes—where the orange line represents the mean and the light orange region represents the entire range of values for the seven sulfur assimilation genes. Met4 complex cofactor MET32's disorganization is highest at 7.5 min while the sulfur assimilation genes and MET30, both of which are activated by the Met4 complex, reach their greatest nucleosome disorganization between 15–30 min.

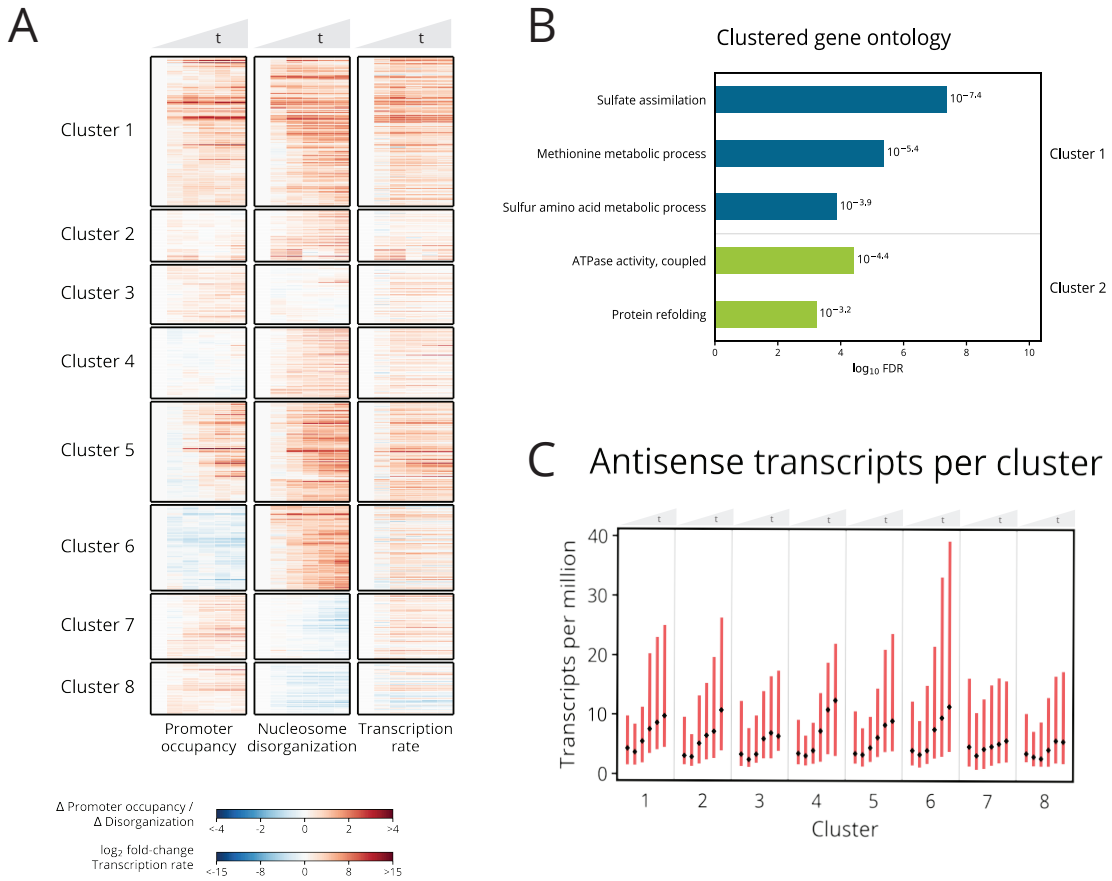

Figure 6. Small fragment promoter occupancy enrichment and gene body nucleosome disorganization reveal stress response pathway timing and patterns with antisense transcription . **(A)** Hierarchical clustering of 832 genes from the union of the 500 greatest increase in average promoter occupancy or 500 greatest increase in average nucleosome disorganization. Clusters 6–8 contain genes with anti-correlated chromatin dynamics. **(B)** GO enrichment analysis identifies clusters 1 and 2 for sulfur and protein refolding pathways respectively. **(C)** Antisense transcript level for each cluster across the time course. Cluster 6 shows enrichment of antisense transcripts matching increased nucleosome disorganization and decreased in promoter occupancy in A.

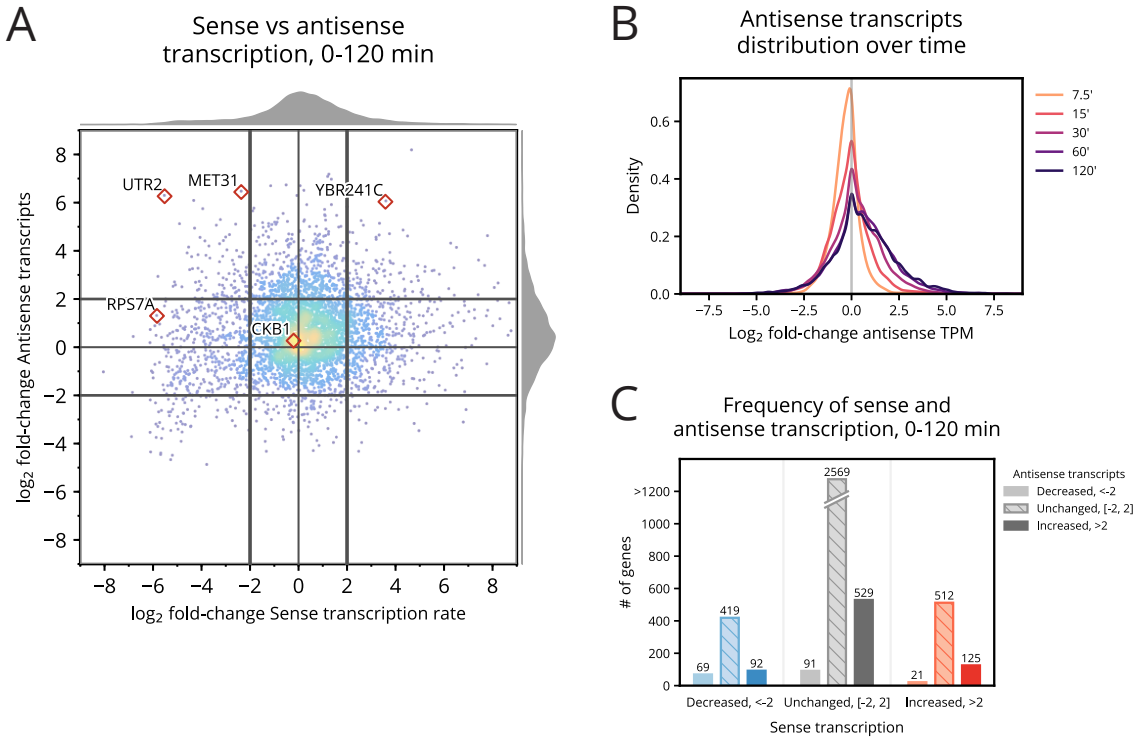

Figure 7. Cadmium induces changes in both sense and antisense transcription. **(A)** Distribution of the $\log_2$ fold-change in sense transcription against the $\log_2$ fold-change in antisense transcripts from 0–120 min. Antisense transcripts are enriched genome-wide by 120 min. **(B)** Distribution of the $\log_2$ fold-change in antisense transcripts for each time point following 0 min. Antisense transcripts monotonically increase through the 120-min. **(C)** Counts of genes that exhibit decreased, unchanged, and increased sense and antisense transcripts from 0 and 120 minutes. Unchanging and increased sense transcription exhibit positively skewed enrichment of antisense transcripts.

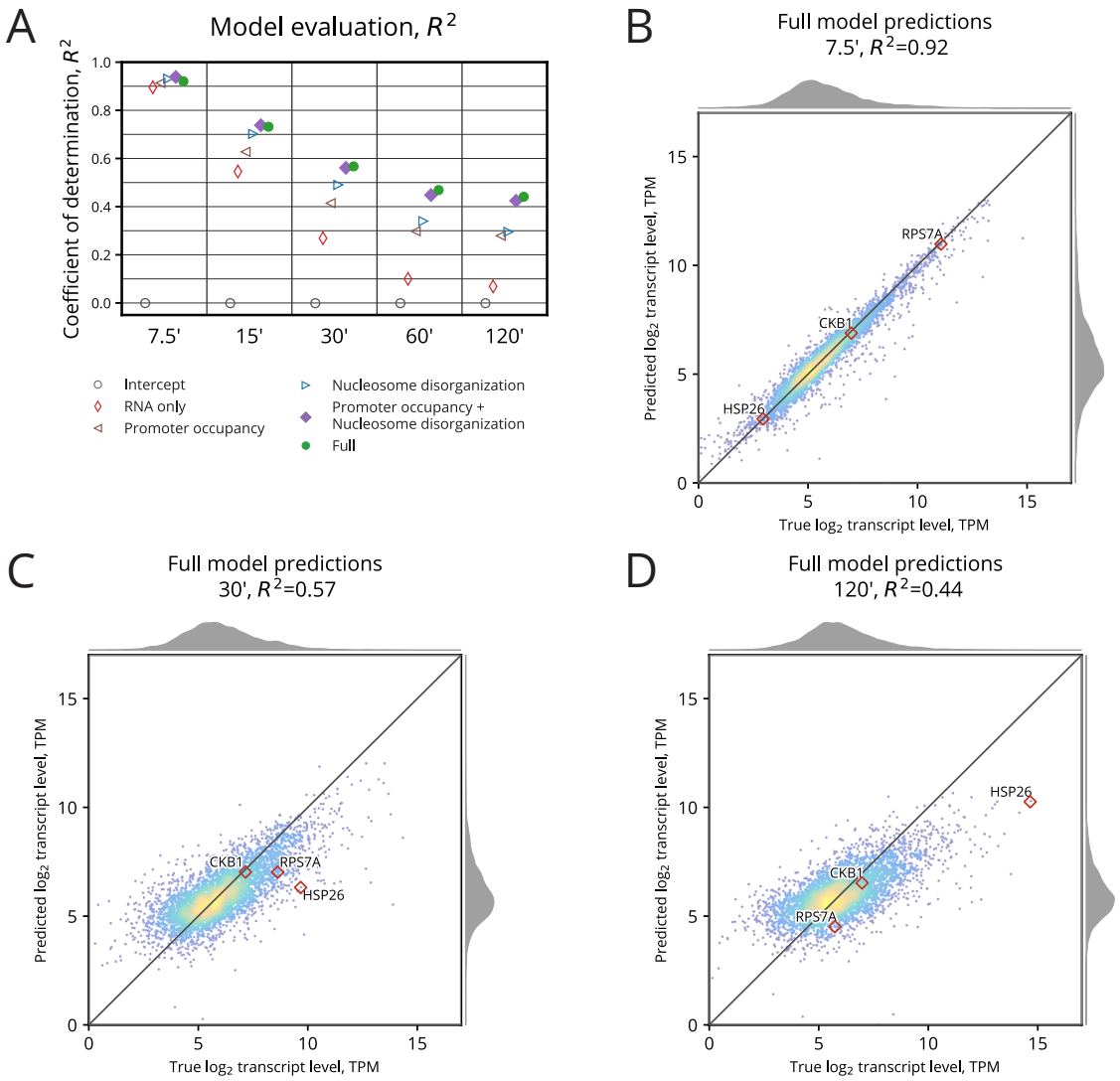

Figure 8. Chromatin occupancy dynamics are predictive of gene expression. **(A)** Comparison of each GP model's performance using its coefficient of determination, R^2. The Full model incorporating all chromatin features and 0 min transcript level outperforms all other models for 30–120 min. Later time points rely less on 0 min transcript level for prediction, so the marginal gain in statistical power between features becomes more evident. **(B)** Comparison between true and predicted $\log_2$ transcript level for the Full model after 7.5 min. Most genes are well predicted using 0 min transcript level. **(C)** Full model predictions at 30 min. Predictions remain well correlated, but less than at 7.5 min. **(D)** Full model predictions at 120 min. After two full hours have elapsed, transcript level predictions have become a bit less correlated, but still, R^2 remains 0.44.

References

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

Barbey R, Baudouin-Cornu P, Lee TA, Rouillon A, Zarzov P, Tyers M, Thomas D. 2005. Inducible dissociation of SCF(Met30) ubiquitin ligase mediates a rapid transcriptional response to cadmium. *EMBO J.* **24**: 521–532.

Belsky JA, MacAlpine HK, Lubelsky Y, Hartemink AJ, MacAlpine DM. 2015. Genome-wide chromatin footprinting reveals changes in replication origin architecture induced by pre-RC assembly. *Genes Dev.* **29**: 212–224.

Benesch JLP, Aquilina JA, Baldwin AJ, Rekas A, Stengel F, Lindner RA, Basha E, Devlin GL, Horwitz J, Vierling E, Carver JA, Robinson CV. 2010. The quaternary organization and dynamics of the molecular chaperone HSP26 are thermally regulated. *Chem. Biol.* **17**: 1008–1017.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**: 289–300.

Blaiseau PL, Thomas D. 1998. Multiple transcriptional activation complexes tether the yeast activator Met4 to DNA. *EMBO J.* **17**: 6327–6336.

Blaiseau PL, Isnard AD, Surdin-Kerjan Y, Thomas D. 1997. Met31p and Met32p, two related zinc finger proteins, are involved in transcriptional regulation of yeast sulfur amino acid metabolism. *Mol. Cell. Biol.* **17**: 3640–3648.

Boy-Marcotte E, Lagniel G, Perrot M, Bussereau F, Boudsocq A, Jacquet M, Labarre J. 1999. The heat shock response in yeast: Differential regulations and contributions of the Msn2p/Msn4p and Hsf1p regulons. *Mol. Microbiol.* **33**: 274–283.

Brahma S, Henikoff S. 2019. RSC-associated subnucleosomes define MNase-sensitive promoters in yeast. *Mol. Cell* **73**: 238–249.e3.

Brennan RJ, Schiestl RH. 1996. Cadmium is an inducer of oxidative stress in yeast. *Mutat. Res.* **356**: 171–178.

Brogaard K, Xi L, Wang JP, Widom J. 2012. A map of nucleosome positions in yeast at base-pair resolution. *Nature* **486**: 496–501.

Camblong J, Iglesias N, Fickentscher C, Dieppois G, Stutz F. 2007. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. *Cell* **131**: 706–717.

Carrillo E, Ben-Ari G, Wildenhain J, Tyers M, Grammentz D, Lee TA. 2012. Characterizing the roles of Met31 and Met32 in coordinating Met4-activated transcription in the absence of Met30. *Mol. Biol. Cell* **23**: 1928–1942.

Cashikar AG, Duennwald M, Lindquist SL. 2005. A chaperone pathway in protein disaggregation: Hsp26 alters the nature of protein aggregates to facilitate reactivation by Hsp104. *J. Biol. Chem.* **280**: 23869–23875.

Chen J, Pederson DS. 1993. A distal heat shock element promotes the rapid response to heat shock of the HSP26 gene in the yeast Saccharomyces cerevisiae. *J. Biol. Chem.* **268**: 7442–7448.

Chereji RV, Ocampo J, Clark DJ. 2017. MNase-sensitive complexes in yeast: Nucleosomes and non-histone barriers. *Mol. Cell* **65**: 565–577.e3.

605 Cormier L, Barbey R, Kuras L. 2010. Transcriptional plasticity through differential
606 assembly of a multiprotein activation complex. *Nucleic Acids Res.* **38**: 4998–5014.

607 Dormer UH, Westwater J, McLaren NF, Kent NA, Mellor J, Jamieson DJ. 2000.
608 Cadmium-inducible expression of the yeast GSH1 gene requires a functional sulfur-
609 amino acid regulatory network. *J. Biol. Chem.* **275**: 32611–32616.

610 Faller P, Kienzler K, Krieger-Liszkay A. 2005. Mechanism of Cd2+ toxicity: Cd2+
611 inhibits photoactivation of Photosystem II by competitive binding to the essential
612 Ca2+ site. *Biochim. Biophys. Acta* **1706**: 158–164.

613 Fauchon M, Lagniel G, Aude JC, Lombardia L, Soularue P, Petat C, Marguerie G, Sen-
614 tenac A, Werner M, Labarre J. 2002. Sulfur sparing in the yeast proteome in response
615 to sulfur demand. *Mol. Cell* **9**: 713–723.

616 Franzmann TM, Menhorn P, Walter S, Buchner J. 2008. Activation of the chaperone
617 Hsp26 is controlled by the rearrangement of its thermosensor domain. *Mol. Cell* **29**:
618 207–216.

619 Gardarin A, Chédin S, Lagniel G, Aude JC, Godat E, Catty P, Labarre J. 2010. Endo-
620 plasmic reticulum is a major target of cadmium toxicity in yeast. *Mol. Microbiol.* **76**:
621 1034–1048.

622 Geisberg JV, Moqtaderi Z, Fan X, Ozsolak F, Struhl K. 2014. Global analysis of mRNA
623 isoform half-lives reveals stabilizing and destabilizing elements in yeast. *Cell* **156**:
624 812–824.

625 Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif.
626 *Bioinformatics* **27**: 1017–1018.

627 Hartwig A. 2001. Zinc finger proteins as potential targets for toxic metal ions: Differ-
628 ential effects on structure and function. *Antioxid. Redox Signal.* **3**: 625–634.

Henikoff JG, Belsky JA, Krassovsky K, MacAlpine DM, Henikoff S. 2011. Epigenome characterization at single base-pair resolution. *Proc. Natl. Acad. Sci. U. S. A.* **108**: 18318–18323.

Hosiner D, Gerber S, Lichtenberg-Fraté H, Glaser W, Schüller C, Klipp E. 2014. Impact of acute metal stress in Saccharomyces cerevisiae. *PLoS One* **9**: e83330.

Huisinga KL, Pugh BF. 2004. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. *Mol. Cell* **13**: 573–585.

Jin YH, Clark AB, Slebos RJC, Al-Refai H, Taylor JA, Kunkel TA, Resnick MA, Gordenin DA. 2003. Cadmium is a mutagen that acts by inhibiting mismatch repair. *Nat. Genet.* **34**: 326–329.

Jin YH, Dunlap PE, McBride SJ, Al-Refai H, Bushel PR, Freedman JH. 2008. Global transcriptome and deletome profiles of yeast exposed to transition metals. *PLoS Genet.* **4**: e1000053.

Kaiser P, Flick K, Wittenberg C, Reed SI. 2000. Regulation of transcription by ubiquitination without proteolysis: Cdc34/SCF(Met30)-mediated inactivation of the transcription factor Met4. *Cell* **102**: 303–314.

Kawahata M, Masaki K, Fujii T, Iefuji H. 2006. Yeast genes involved in response to lactic acid and acetic acid: Acidic conditions caused by the organic acids in Saccharomyces cerevisiae cultures induce expression of intracellular metal metabolism genes regulated by Aft1p. *FEMS Yeast Res.* **6**: 924–936.

Kim TS, Liu CL, Yassour M, Holik J, Friedman N, Buratowski S, Rando OJ. 2010. RNA polymerase mapping during stress responses reveals widespread nonproductive transcription in yeast. *Genome Biol.* **11**: R75.

Klopfenstein DV, Zhang L, Pedersen BS, Ramírez F, Warwick Vesztrocy A, Naldi A, Mungall CJ, Yunes JM, Botvinnik O, Weigel M, Dampier W, Dessimoz C, Flick P, Tang H. 2018. GOATOOLS: A Python library for Gene Ontology analyses. *Sci. Rep.* **8**: 10872.

Kornienko AE, Guenzl PM, Barlow DP, Pauler FM. 2013. Gene regulation by the act of long non-coding RNA transcription. *BMC Biol.* **11**: 59.

Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrín-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B, Richards DP, Canadien V, Lalev A, Mena F, Wong P, Starostine A, Canete MM, Vlasblom J, Wu S, Orsi C, Collins SR, Chandran S, Haw R, Rilstone JJ, Gandi K, Thompson NJ, Musso G, St Onge P, Ghanny S, Lam MHY, Butland G, Altaf-Ul AM, Kanaya S, Shilatifard A, O'Shea E, Weissman JS, Ingles CJ, Hughes TR, Parkinson J, Gerstein M, Wodak SJ, Emili A, Greenblatt JF. 2006. Global landscape of protein complexes in the yeast saccharomyces cerevisiae. *Nature* **440**: 637–643.

Kubik S, Bruzzone MJ, Albert B, Shore D. 2017. A reply to "MNase-sensitive complexes in yeast: Nucleosomes and non-histone barriers," by Chereji et al. *Mol. Cell* **65**: 578–580.

Kulaeva OI, Hsieh FK, Studitsky VM. 2010. RNA polymerase complexes cooperate to relieve the nucleosomal barrier and evict histones. *Proc. Natl. Acad. Sci. U. S. A.* **107**: 11325–11330.

Kuras L, Cherest H, Surdin-Kerjan Y, Thomas D. 1996. A heteromeric complex containing the centromere binding factor 1 and two basic leucine zipper factors, Met4 and Met28, mediates the transcription activation of yeast sulfur metabolism. *EMBO J.* **15**: 2519–2529.

Kuras L, Rouillon A, Lee T, Barbey R, Tyers M, Thomas D. 2002. Dual regulation of the Met4 transcription factor by ubiquitin-dependent degradation and inhibition of promoter recruitment. *Mol. Cell* **10**: 69–80.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**: R25.

Lee CK, Shibata Y, Rao B, Strahl BD, Lieb JD. 2004. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat. Genet.* **36**: 900–905.

Lenstra TL, Benschop JJ, Kim T, Schulze JM, Brabers NACH, Margaritis T, van de Pasch LAL, van Heesch SAAC, Brok MO, Groot Koerkamp MJA, Ko CW, van Leenen D, Sameith K, van Hooff SR, Lijnzaad P, Kemmeren P, Hentrich T, Kobor MS, Buratowski S, Holstege FCP. 2011. The specificity and topology of chromatin interaction pathways in yeast. *Mol. Cell* **42**: 536–549.

Lubliner S, Keren L, Segal E. 2013. Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. *Nucleic Acids Res.* **41**: 5569–5581.

MacIsaac KD, Wang T, Gordon DB, Gifford DK, Stormo GD, Fraenkel E. 2006. An improved map of conserved regulatory sites for Saccharomyces cerevisiae. *BMC Bioinformatics* **7**: 113.

McIsaac RS, Petti AA, Bussemaker HJ, Botstein D. 2012. Perturbation-based analysis and modeling of combinatorial regulation in the yeast sulfur assimilation pathway. *Mol. Biol. Cell* **23**: 2993–3007.

Miller C, Schwalb B, Maier K, Schulz D, Dümcke S, Zacher B, Mayer A, Sydow J, Marcinowski L, Dölken L, Martin DE, Tresch A, Cramer P. 2011. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol. Syst. Biol.* **7**: 458.

Miller KA, DiDone L, Krysan DJ. 2010. Extracellular secretion of overexpressed glycosylphosphatidylinositol-linked cell wall protein Utr2/Crh2p as a novel protein quality control mechanism in saccharomyces cerevisiae. *Eukaryot. Cell* **9**: 1669–1679.

Momose Y, Iwahashi H. 2001. Bioassay of cadmium using a DNA microarray: Genome-wide expression patterns of Saccharomyces cerevisiae response to cadmium. *Environ. Toxicol. Chem.* **20**: 2353–2360.

Nadal-Ribelles M, Solé C, Xu Z, Steinmetz LM, de Nadal E, Posas F. 2014. Control of Cdc28 CDK1 by a stress-induced lncRNA. *Mol. Cell* **53**: 549–561.

Nocetti N, Whitehouse I. 2016. Nucleosome repositioning underlies dynamic gene expression. *Genes Dev.* **30**: 660–672.

Ouni I, Flick K, Kaiser P. 2010. A transcriptional activator is part of an SCF ubiquitin ligase to control degradation of its cofactors. *Mol. Cell* **40**: 954–964.

Park D, Morris AR, Battenhouse A, Iyer VR. 2014. Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements. *Nucleic Acids Res.* **42**: 3736–3749.

Patton EE, Peyraud C, Rouillon A, Surdin-Kerjan Y, Tyers M, Thomas D. 2000. SCF(Met30)-mediated control of the transcriptional activator Met4 is required for the G(1)-S transition. *EMBO J.* **19**: 1613–1624.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**: 2825–2830.

Pereira Y, Lagniel G, Godat E, Baudouin-Cornu P, Junot C, Labarre J. 2008. Chromate causes sulfur starvation in yeast. *Toxicol. Sci.* **106**: 400–412.

Petti AA, McIsaac RS, Ho-Shing O, Bussemaker HJ, Botstein D. 2012. Combinatorial control of diverse metabolic and physiological functions by transcriptional regulators of the yeast sulfur assimilation pathway. *Mol. Biol. Cell* **23**: 3008–3024.

Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, Olson S, Weinberg D, Baker KE, Graveley BR, Coller J. 2015. Codon optimality is a major determinant of mRNA stability. *Cell* **160**: 1111–1124.

Rabani M, Levin JZ, Fan L, Adiconis X, Raychowdhury R, Garber M, Gnirke A, Nusbaum C, Hacohen N, Friedman N, Amit I, Regev A. 2011. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat. Biotechnol.* **29**: 436–442.

Ramachandran S, Ahmad K, Henikoff S. 2017. Transcription and remodeling produce asymmetrically unwrapped nucleosomal intermediates. *Mol. Cell* **68**: 1038–1053.e4.

Reja R, Vinayachandran V, Ghosh S, Pugh BF. 2015. Molecular mechanisms of ribosomal protein gene coregulation. *Genes Dev.* **29**: 1942–1954.

Rhee HS, Pugh BF. 2012. Genome-wide structure and organization of eukaryotic preinitiation complexes. *Nature* **483**: 295–301.

Schwabish MA, Struhl K. 2004. Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II. *Mol. Cell. Biol.* **24**: 10111–10117.

Sharma SK, Goloubinoff P, Christen P. 2008. Heavy metal ions are potent inhibitors of protein folding. *Biochem. Biophys. Res. Commun.* **372**: 341–345.

Shivaswamy S, Iyer VR. 2008. Stress-dependent dynamics of global chromatin remodeling in yeast: Dual role for SWI/SNF in the heat shock stress response. *Mol. Cell. Biol.* **28**: 2221–2234.

Smale ST, Kadonaga JT. 2003. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* **72**: 449–479.

Susek RE, Lindquist S. 1990. Transcriptional derepression of the Saccharomyces cerevisiae HSP26 gene during heat shock. *Mol. Cell. Biol.* **10**: 6362–6373.

Swamy KBS, Lin CH, Yen MR, Wang CY, Wang D. 2014. Examining the condition-specific antisense transcription in S. cerevisiae and S. paradoxus. *BMC Genomics* **15**: 521.

Teves SS, Henikoff S. 2011. Heat shock reduces stalled RNA polymerase II and nucleosome turnover genome-wide. *Genes Dev.* **25**: 2387–2397.

The Gene Ontology Consortium. 2019. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**: D330–D338.

Till P, Mach RL, Mach-Aigner AR. 2018. A current view on long noncoding RNAs in yeast and filamentous fungi. *Appl. Microbiol. Biotechnol.* **102**: 7319–7331.

Toesca I, Nery CR, Fernandez CF, Sayani S, Chanfreau GF. 2011. Cryptic transcription mediates repression of subtelomeric metal homeostasis genes. *PLoS Genet.* **7**: e1002163.

Treger JM, Schmitt AP, Simon JR, McEntee K. 1998. Transcriptional factor mutations reveal regulatory complexities of heat shock and newly identified stress genes in Saccharomyces cerevisiae. *J. Biol. Chem.* **273**: 26875–26879.

Tripuraneni V, Memisoglu G, Zhu W, Tran T, Hartemink AJ, Haber JE, MacAlpine DM. 2019. Local nucleosome dynamics and eviction following a double-strand break are

774 reversible by NHEJ-mediated repair in the absence of DNA replication. *bioRxiv* page
775 866673.

776 Vance KW, Ponting CP. 2014. Transcriptional regulatory functions of nuclear long non-
777 coding RNAs. *Trends Genet.* **30**: 348–355.

778 Venters BJ, Wachi S, Mavrich TN, Andersen BE, Jena P, Sinnamon AJ, Jain P, Rolleri NS,
779 Jiang C, Hemeryck-Walsh C, Pugh BF. 2011. A comprehensive genomic binding map
780 of gene and chromatin regulatory proteins in Saccharomyces. *Mol. Cell* **41**: 480–492.

781 Vinayachandran V, Reja R, Rossi MJ, Park B, Rieber L, Mittal C, Mahony S, Pugh BF.
782 2018. Widespread and precise reprogramming of yeast protein-genome interactions
783 in response to heat shock. *Genome Res.* **28**: 357–366.

784 Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski
785 E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Jarrod
786 Millman K, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey C, Polat İ,
787 Feng Y, Moore EW, Vand erPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I,
788 Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P,
789 Contributors S. 2020. SciPy 1.0: Fundamental algorithms for scientific computing in
790 Python. *Nature Methods* **17**: 261–272.

791 Wagner GP, Kin K, Lynch VJ. 2012. Measurement of mRNA abundance using RNA-seq
792 data: RPKM measure is inconsistent among samples. *Theory Biosci.* **131**: 281–285.

793 Ward JH. 1963. Hierarchical grouping to optimize an objective function. *J. Am. Stat.*
794 *Assoc.* **58**: 236–244.

795 Weiner A, Chen HV, Liu CL, Rahat A, Klien A, Soares L, Gudipati M, Pfeffner J, Regev
796 A, Buratowski S, Pleiss JA, Friedman N, Rando OJ. 2012. Systematic dissection of
797 roles for chromatin regulators in a yeast stress response. *PLoS Biol.* **10**: e1001369.

Weiner A, Hsieh THS, Appleboim A, Chen HV, Rahat A, Amit I, Rando OJ, Friedman N. 2015. High-resolution chromatin dynamics during a yeast stress response. *Mol. Cell* **58**: 371–386.

Wiederhold E, Gandhi T, Permentier HP, Breitling R, Poolman B, Slotboom DJ. 2009. The yeast vacuolar membrane proteome. *Mol. Cell. Proteomics* **8**: 380–392.

Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bähler J. 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453**: 1239–1243.

Yang E, van Nimwegen E, Zavolan M, Rajewsky N, Schroeder M, Magnasco M, Darnell Jr JE. 2003. Decay rates of human mRNAs: Correlation with functional characteristics and sequence attributes. *Genome Res.* **13**: 1863–1872.