Linking the dynamics of chromatin occupancy and transcription with predictive models

Trung Q. Tran¹, Heather K. MacAlpine², Vinay Tripuraneni², Sneha Mitra¹,
David M. MacAlpine^{2,3,*}, and Alexander J. Hartemink^{1,3,*}

¹Department of Computer Science,

Duke University, Durham, NC 27708, USA

²Department of Pharmacology and Cancer Biology,

Duke University Medical Center, Durham, NC 27710, USA

³Center 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

December 14, 2020

Abstract

Though the sequence of the genome within each eukaryotic cell is essentially fixed, it exists within 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 changes and how those changes 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.

20 Introduction

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

29 Numerous studies have made major strides in characterizing the roles of protein
 30 complexes involved in transcription. Chromatin immunoprecipitation (ChIP) has been
 31 used to assay binding sites of hundreds of proteins on a genomic scale, including factors
 32 involved in SAGA-dominated stress-related pathways and TFIID-dominated housekeep-
 33 ing pathways (Venters *et al.* 2011). Likewise, studies have probed proteins involved in
 34 the formation of the pre-initiation complex required for transcription initiation (Rhee
 35 and Pugh 2012). The role of numerous chromatin remodelers and their interactions
 36 have been characterized in detail through ChIP, proteomics, and gene expression anal-
 37 ysis of deletion mutants (Krogan *et al.* 2006; Lenstra *et al.* 2011; Mavrich *et al.* 2008;
 38 Shivaswamy and Iyer 2008; Weiner *et al.* 2012, 2015). However, limitations in these
 39 methods, including lack of antibodies for ChIP or viability of deletion strains, are of-
 40 ten constraining. Analysis can be complicated by the difficulty in disentangling direct
 41 chromatin effects from the pleiotropic action of the many factors and remodelers that
 42 impinge upon transcription, often indirectly. These and other issues contribute to our
 43 still limited understanding of the dynamic interplay of the chromatin landscape and
 44 transcription.

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 at nucleotide resolution with transcriptional regulation, especially regulation induced by environmental perturbations. Here, we utilize a high-resolution spatiotemporal stress response data set to elucidate the relationship between chromatin organization and gene expression by developing general strategies and models to analyze, genome-wide, chromatin dynamics relative to changes in transcription.

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

70 proteins. A nucleotide resolution view of chromatin occupancy dynamics in response
71 to cadmium stress would allow us to associate and infer relationships between these
72 chromatin changes and those in transcription. Yeast cells were exposed to cadmium
73 and samples were collected over a two-hour time course (Fig. 1A). Chromatin occu-
74 pancy and positioning dynamics were profiled using paired-end MNase-seq to map
75 DNA-binding factors at base pair resolution (Fig. 1B). Concurrently, transcripts were
76 interrogated using strand-specific total RNA-seq (Fig. 1C).

77 To evaluate our data and methods, we considered the well-studied stress response
78 gene *HSP26*, whose role is to facilitate the disaggregation of misfolded proteins (Cashikar
79 *et al.* 2005). Hsp26 has been implicated in responses to many stress conditions, includ-
80 ing heat shock (Benesch *et al.* 2010; Franzmann *et al.* 2008), acidity (Kawahata *et al.*
81 2006), sulfur starvation (Pereira *et al.* 2008), and metal toxicity (Hosiner *et al.* 2014;
82 Momose and Iwahashi 2001). Furthermore, several transcription factors, including
83 Hsf1, Met4, and Met32, have been found to bind in the well-characterized promoter of
84 *HSP26* (Boy-Marcotte *et al.* 1999; Carrillo *et al.* 2012; Chen and Pederson 1993; Susek
85 and Lindquist 1990; Treger *et al.* 1998). Given this context, *HSP26* serves as a useful
86 test case because we understand many aspects of its local chromatin dynamics when it
87 is activated under stress conditions.

88 We observed significant changes in the chromatin around the transcription start site
89 (TSS) of *HSP26* (Fig. 2A), coinciding with a dramatic increase in its transcript level.
90 Upstream, in the promoter of *HSP26*, nucleosome-sized fragments of length 144–174
91 bp are replaced by small fragments less than 100 bp. In the gene body of *HSP26*,
92 nucleosome-sized fragments become “fuzzy”, increasing in positional and fragment-
93 length variability (Fig. 2A). Nucleosomes upstream of *HSP26* are known to be evicted
94 (Lee *et al.* 2004) and replaced by smaller factors associated with transcription initia-
95 tion, pushing gene body nucleosomes downstream (Fig. 2B,C). Then, active transcrip-

tion by RNA polymerases displaces and evicts nucleosomes in its path (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 the gene body of *HSP26*.

To quantify these complex transcription-associated chromatin dynamics genome-wide, we defined two scores for each gene, a “small fragment occupancy” score of small fragments appearing in a gene’s promoter, and a measure of “nucleosome disorganization” within its gene body using information entropy. Additionally, to account for variations in RNA stability, we estimated transcription rates from our measured transcript levels using published mRNA decay rates (Geisberg *et al.* 2014; Miller *et al.* 2011; Presnyak *et al.* 2015).

Using these measures, we are able to succinctly describe relationships between chromatin dynamics and transcription in a range of genes, from activated *HSP26* (Fig. 2D), to repressed *RPS7A* (Supplemental Fig. 1), to unchanging *CKB1* (Supplemental Fig. 2). Averaging these two measures of the chromatin across the time course, and then ranking all genes by the resulting “combined chromatin” score, we observed large-scale coordination between chromatin and transcription across a significant proportion of the genome (Fig. 3A,B).

Globally, log fold-changes in transcription show a significant positive Pearson correlation with changes in each of our chromatin measures: 0.49 for small fragment occupancy (Fig. 3C), 0.61 for nucleosome disorganization (Fig. 3D), and 0.68 for combined chromatin (Supplemental Fig. 3A). The high correlation between combined chromatin and transcription, along with a lower 0.33 correlation between small fragment occupancy and nucleosome disorganization (Supplemental Fig. 3B), suggests that small fragment occupancy and nucleosome disorganization each provide orthogonal statistical power in describing changes in the chromatin relative to changes in transcription.

121 *Changes in nucleosome and small factor occupancy at TSSs recapitulate genome-wide tran-*
122 *scriptional response to cadmium*

123 To determine whether chromatin dynamics alone could recapitulate known response
124 to cadmium exposure, we performed Gene Ontology (GO) enrichment analysis of the
125 300 genes with the highest and lowest values for each chromatin measure. We identi-
126 fied, with varying levels of false discovery rate (FDR) significance, regulation pathways
127 implicated under cadmium exposure. We further validated these chromatin-identified
128 pathways using literature and a separate GO enrichment analysis based on changes in
129 transcription (Supplemental Tables S1 and S2).

130 One of the established responses for cells undergoing stress involves shutting down
131 ribosomal and other translation-related pathways (Hosiner *et al.* 2014; Reja *et al.* 2015;
132 Vinayachandran *et al.* 2018). Using our simple chromatin measures, ribosomal and
133 translation-related GO terms emerged as the most significantly down-regulated, with
134 FDR values often much lower than 10^{-10} (Fig. 4A).

135 Translation-related genes are repressed as a tightly regulated cluster, but pathways
136 activated under cadmium exposure are also recovered as the most significantly up-
137 regulated by our chromatin scores, albeit with FDR values above 10^{-4} (Fig. 4B). Con-
138 sistent with previous cadmium and heavy metal stress response studies (Faller *et al.*
139 2005; Fauchon *et al.* 2002; Hartwig 2001) and our own transcriptional GO enrichment
140 analysis (Supplemental Table S2), two major cadmium-response pathways were im-
141 plicated by changes in the chromatin: sulfur assimilation and protein folding. While
142 small fragment occupancy identified sulfate assimilation and stress response terms with
143 the greatest significance (FDR of $10^{-3.9}$), nucleosome disorganization was required to
144 identify protein refolding and sulfur amino acid metabolic process terms. Our differ-
145 ent measures computed from chromatin are sufficient to accurately recover high-level
146 stress response pathways induced and repressed by cadmium exposure.

147 *High-resolution time course recovers cascading induction of sulfur pathways*

148 Because of the significant involvement of sulfur assimilation in the cell's response to
 149 cadmium, we next sought to detail changes in the chromatin related to the activation
 150 of sulfur pathways. The heavy demand for sulfur arises because it is required for the
 151 biosynthesis of the cadmium-chelating glutathione (Fauchon *et al.* 2002). Sulfur path-
 152 ways are activated through Met4 and its binding complex, comprised of cis-binding fac-
 153 tors Cbf1 and Met31/Met32, and accessory factor Met28 (Blaiseau and Thomas 1998;
 154 Kuras *et al.* 1996). Met4 is negatively regulated through ubiquitination by SCF^{Met30}
 155 (Barbey *et al.* 2005; Kaiser *et al.* 2000; Kuras *et al.* 2002)(Fig. 5A). In our study, we
 156 identified novel features of the chromatin in the cascading events that regulate the sul-
 157 fur metabolic pathways (Fig. 5B): (i) the activation of the Met4 complex through its
 158 cofactors, (ii) the activation of the sulfur pathways by Met4, and (iii) the subsequent
 159 down-regulation of Met4 activity by SCF^{Met30}, evident in diminished transcription of
 160 Met4-regulated genes.

161 Upon deubiquitination by cadmium (Barbey *et al.* 2005), Met4 becomes function-
 162 ally active and induces its own cofactors (Barbey *et al.* 2005; McIsaac *et al.* 2012) and,
 163 through feedforward regulation between Met4 and Met32, activates sulfur pathway
 164 genes (Carrillo *et al.* 2012; McIsaac *et al.* 2012). We observed this activation not only
 165 in increased transcription within 7.5 minutes for *MET32* and *MET28*, but also in dra-
 166 matic nucleosome disorganization of *MET32* (Supplemental Fig. 4) and increased small
 167 fragment occupancy for *MET28*.

168 While Met31 shares a binding motif and largely overlaps in function with Met32
 169 (Blaiseau *et al.* 1997), it is not as prominent as Met32 in the activation of sulfur path-
 170 ways (Carrillo *et al.* 2012; McIsaac *et al.* 2012; Petti *et al.* 2012). In response to cad-
 171 mium, the transcription of *MET31* is repressed, but the chromatin around the gene ex-
 172 hibits an unexpected behavior in light of this: although *MET31* expression is repressed,

its nucleosomes become highly disorganized. Leveraging our stranded RNA-seq data, we noticed significantly increasing antisense transcription over the time course (Supplemental Fig. 5). Additionally, downstream of the transcription end site (TES) of *MET31*, small fragments become enriched at a Met31/Met32 binding motif. Taken together, our data suggests that *MET31* is being regulated by non-coding RNA (ncRNA) antisense transcription.

Following activation of the Met4 complex (Carrillo *et al.* 2012; McIsaac *et al.* 2012), small fragment occupancy, nucleosome disorganization, and transcription increase for the seven sulfur assimilation genes (Fig. 5C) and many downstream genes within 15 minutes. Additionally, the Met4 complex induces a sulfur-sparing transcriptional-switch between functionally similar isoenzymes to indirectly contribute sulfur required for chelation. This switch includes replacing sulfur-rich Pdc1 with sulfur-lacking Pdc6, Ald6 with Ald4, and Eno2 with Eno1 (Fauchon *et al.* 2002). We see evidence of these substitutions between isoenzyme pairs in our data, with the most dramatic changes evident in the small fragment occupancy of *PDC6* (details in Supplemental Fig. 6) and *PDC1*.

Following induction of the sulfur pathways, the activating roles of Met32 and Met4 diminish upon regulation by SCF^{Met30} (Ouni *et al.* 2010; Patton *et al.* 2000). This regulation is observed in our data in the gradually increasing transcription and nucleosome disorganization of *MET30* throughout the time course, as well as in how the nucleosome disorganization scores of *MET32* and many of the sulfur assimilation genes gradually diminish after an early peak (Fig. 5B,D).

Together, these results and analyses complement established transcriptional and ChIP-based studies by detailing chromatin dynamics of the sulfur metabolic pathways and identifying a potentially novel regulatory mechanism for *MET31* through antisense transcription.

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

201 We selected the 500 genes exhibiting the greatest average increase in either small frag-
202 ment occupancy or nucleosome disorganization, and performed hierarchical clustering
203 on the resulting 832 genes (fewer than 1000 because many were in both sets). Cluster-
204 ing revealed distinct temporal patterns in small fragment occupancy and nucleosome
205 disorganization among the genes (Fig. 6A). GO enrichment analysis identified differ-
206 ent stress response pathways in two of the clusters (Fig. 6B), suggesting that chromatin
207 changes in these pathways differ in their temporal pattern. Clusters 6–8 reveal unex-
208 pected anti-correlated relationships between chromatin and transcription for genes in
209 these clusters. For genes in cluster 6, some of the anti-correlation can be attributed
210 to antisense transcription (Fig. 6C), as previously highlighted in *MET31*. But in clus-
211 ter 7, *MCD4*, which codes for an endoplasmic reticulum membrane protein, counter-
212 intuitively exhibits chromatin with nucleosomes that become more organized despite
213 increased sense and no evident antisense transcription (Supplemental Fig. 7).

214 Genome-wide, we observed that antisense transcription manifests itself with min-
215 imal apparent connection to sense transcription (Fig. 7A). Nevertheless, we did de-
216 tect two general phenomena, each consistent with prior studies. First, as also seen in
217 other environmental conditions (Kim *et al.* 2010; Till *et al.* 2018; Wilhelm *et al.* 2008),
218 yeast undergoing cadmium stress induce pervasive antisense transcription. As the time
219 course progresses, more and more genes exhibit increased levels of antisense transcrip-
220 tion (Fig. 7B). Even among the 3,199 genes whose sense transcription changes only
221 minimally, 542 exhibit at least a four-fold increase in antisense transcription (Fig. 7C).
222 Second, previous studies have found antisense transcription can be associated with ei-
223 ther repression or activation of target genes (Kornienko *et al.* 2013; Swamy *et al.* 2014;
224 Till *et al.* 2018; Vance and Ponting 2014), and we observed the same phenomenon. Un-

der cadmium stress, we identified 200 genes whose antisense transcripts increased at least four-fold and whose sense transcripts changed by at least four-fold. Among those, 104 had repressed sense transcription—*e.g.*, *MET31* and *UTR2*, whose overexpression has been linked with endoplasmic reticulum stress (Miller *et al.* 2010) (Supplemental Fig. 8)—but 96 had activated sense transcription, including the gene *YBR241C* (Supplemental Fig. 9), coding for a vacuole localization protein (Wiederhold *et al.* 2009).

Motif analysis identifies per-bp binding dynamics of transcription factors

To explore small fragment occupancy more closely, we identified peaks in the signal and quantified the change in binding at each peak over 60 minutes (Fig. 8A). We ran the motif finder FIMO (Grant *et al.* 2011) near peak locations to associate peaks with TFs, and then for each TF, computed its average change in binding occupancy (Fig. 8B). TFs exhibiting the greatest average increase in occupancy include not only the sulfur pathway activators Met4 and Met32, general stress regulators Msn2 and Msn4, and glycolytic activators Gcr1 and Gcr2, but also the iron homeostasis regulators Aft1 and Aft2. Genes with the greatest increase in both Aft1 and Aft2 binding include *SER33*, *LEE1*, and *ENB1*.

For *SER33*, a gene involved in Ser and Gly biosynthesis (Albers *et al.* 2003), we see evidence of Aft1/Aft2 binding near Gcr2 in the promoter (Fig. 8C). Whereas Gcr2 is known to interact with Gcr1, a known regulator of *SER33* (Hu *et al.* 2007), Aft1 and Aft2 have yet to be identified as regulators for *SER33* (Fig. 8D). Additionally, we see enrichment of small fragments near the motifs for known regulators Met32 and Met4, previously identified through ChIP (Carrillo *et al.* 2012). Similarly strong evidence of Aft1/Aft2 binding is found in the promoters of *LEE1* (Supplemental Fig. 10A), a zinc-finger of unknown function, and *ENB1* (Supplemental Fig. 10B), a ferric enterobactin transmembrane transporter (Heymann *et al.* 2000). *ENB1* has only been identified to

be regulated by Aft1 through microarrays (Hu *et al.* 2007). While the iron homeostasis pathways have been previously implicated in heavy metal stress conditions (Halimaa *et al.* 2019; Hosiner *et al.* 2014), our analysis further elucidates the binding dynamics of regulators Aft1 and Aft2 under cadmium stress and, more generally, demonstrates the richness of small fragment signals in MNase-seq data.

Chromatin occupancy changes are predictive of changes in gene expression

Finally, we sought to develop a model to quantify the relationship between our measures of chromatin dynamics and changes in transcription. We used Gaussian process regression models to predict the transcription at each time point based solely on chromatin dynamics and initial transcript levels (at 0 min, before cadmium treatment). We constructed four models to evaluate the inclusion of various measures of the chromatin, culminating in a “full” model that incorporates additional occupancy measures, nucleosome positional shifts (Supplemental Fig. 11), and chromatin measures relative to called antisense transcripts (Supplemental Fig. 12).

Under 10-fold cross-validation, we evaluated each model using the coefficient of determination (R^2), as the proportion of variance each model is able to explain (Fig. 9). For each feature-containing model, prediction performance gradually worsens through the time course as genes’ transcript levels increasingly diverge from their initial values. However, models that include chromatin features consistently outperform a model that just uses initial transcript levels (RNA only), with the gap growing over time. Nucleosome disorganization is more informative than small fragment occupancy, especially at intermediate times; consistent with our other results, combining both measures provides more predictive power than either alone. The full model does not add much to this combination at 7.5 and 15 minutes because early predictions are mainly driven by initial transcript levels. However, by 30 minutes, it begins to outperform all other

models, maintaining an R^2 of 0.44 even two hours after the cell's exposure to cadmium.

While our models cannot ascertain causal links between changes in chromatin and transcription and use measures that do not fully characterize the chromatin state, they nevertheless provide strong evidence that a large proportion of a cell's transcription state can be predicted from simple measures of its chromatin state, even after significant environmental perturbation.

Discussion

In contrast to ChIP-based studies that profile one DNA-binding factor at a time, our study surveys the occupancy of all factors across the entire genome, albeit without explicit information on their identities. While nucleosomes are well-characterized by MNase digestion, profiling the TFs and complexes that regulate gene expression is a more challenging, open problem. Prior work has explored the dynamics of various individual promoter-binding factors including TFs, general TFs, polymerases, mediator, SAGA, TFIID, chromatin remodelers and histone modifications, 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). These studies—along with motif analysis—provide us with useful context to understand the dynamics of small fragments in our MNase-seq data, such as in our characterizations of *HSP26* (Fig. 2), the sulfur pathways (Fig. 5), and the iron homeostasis regulators Aft1/Aft2 (Fig. 8).

Analysis of *MET31*, encoding a Met4 cofactor, revealed chromatin changes linked with increased antisense transcription that may explain how the cell regulates its sense transcription. Moreover, we observed pervasive antisense transcription under cadmium stress, and while this has previously been shown to occur under a variety of environ-

mental perturbations (Camblong *et al.* 2007; Nadal-Ribelles *et al.* 2014; Swamy *et al.* 2014; Toesca *et al.* 2011), we were able to characterize relationships between sense and antisense transcription with regulatory insight from the perspective of the local chromatin landscape. For the 667 genes we identified with antisense transcripts, including chromatin measures relative to those transcripts improved our model (marginally) for predicting sense transcription (Fig. 9A). This benefit can be further explored 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 just the initial transcript level and simple measures of chromatin dynamics, our regression model is able to predict the level of sense transcript with an R^2 at least 0.44, even two hours after cadmium exposure (Fig. 9A). This model can be extended in multiple directions. We can further quantify the chromatin by including additional classes of fragments, by computing new measures of chromatin dynamics, and by considering chromatin beyond 200 bp of a promoter and the first 500 bp of a gene body. Additionally, our data could be modeled with 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 MNase-derived time-series chromatin data.

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.

322 *Growing and sampling cells over the time course*

323 Cells were grown asynchronously in YEPD at 30°C to an OD₆₀₀ of 0.8. Immediately
 324 before the addition of CdCl₂, one sample was removed and cross-linked with formalde-
 325 hyde to a final concentration of 1% for MNase-seq, and another was pelleted and flash
 326 frozen for RNA-seq; these represent time 0. After the addition of CdCl₂ to a final
 327 concentration of 1 mM, samples were taken at 7.5 min, 15 min, 30 min, 60 min, and
 328 120 min, and processed in the above manner, respectively, in preparation for MNase-seq
 329 and RNA-seq. All experiments were repeated independently as biological replicates.

330 *Preparing chromatin*

331 Cells were resuspended with 20 ml of buffer Z (0.56 M sorbitol, 50 mM Tris at pH 7.4)
 332 and 14 μ L of β -ME and 0.5 ml of a 10 mg/mL solution of zymolyase (Sunrise Science
 333 Products) prepared in buffer Z were added. Samples were incubated for 30 min at
 334 24°C with shaking. Cells were centrifuged at 1500 rpm for 6 min at 4°C and then
 335 resuspended in 2.5 ml of NP buffer (1 M sorbitol, 50 mM NaCl, 10 mM Tris at pH 7.4,
 336 5 mM MgCl₂, 1 mM CaCl₂), supplemented with 0.5 mM spermidine, 0.007% β -ME,
 337 and 0.075% NP-40. To determine the best digestion conditions, a four-step titration
 338 of 15 U/ μ L MNase (Worthington) was added to 400 μ L of zymolyase treated cells.
 339 Samples were inverted to mix and digested on the benchtop for 20 min. The reaction
 340 was halted by adding 100 μ L of stop buffer (5% SDS, 50 mM EDTA). Next, proteinase
 341 K was added to a 0.2 mg/mL final concentration, and the samples were inverted and
 342 then incubated overnight at 65°C. DNA was recovered by phenol/chloroform extraction
 343 and isopropanol precipitation.

344 *Preparing MNase sequencing libraries*

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

347 *Preparing RNA sequencing libraries*

348 Illumina sequencing libraries of total RNA were prepared using the Illumina TruSeq
349 Stranded Total RNA Human/Mouse/Rat kit (Cat number RS-122-2201) following the
350 protocol provided by Illumina with Ribo-Zero.

351 *Aligning sequencing reads to the genome*

352 All reads were aligned to the sacCer3/R64 version of the *S. cerevisiae* genome using
353 Bowtie 0.12.7 (Langmead *et al.* 2009).

354 The recovered sequences from all paired-end MNase reads were truncated to 20 bp
355 and aligned in paired-end mode using the following Bowtie parameters: --wrapper
356 basic-0 --time -p 32 -n 2 -l 20 --phred33-quals -m 1 --best --strata -S.

357 The recovered sequences from all single-end RNA reads were truncated to 51 bp
358 and aligned in single-end mode using the same Bowtie parameters.

359 *Processing reads from MNase-seq and RNA-seq replicates*

360 After confirming high concordance between them, MNase-seq and RNA-seq replicates
361 were subsampled and merged to increase read depth and reduce bias from library
362 preparation, sequencing, and digestion. Details of the procedure used to subsample
363 and merge each pair of replicates are provided in Supplemental Method S1. After merg-
364 ing, we had 24,152,389 mapped MNase fragments (pairs of reads) and 42,107,377

mapped RNA reads for each time point.

Selecting a set of genes for analysis

We compiled a set of 4,427 genes for analysis. A gene was chosen if it satisfied five criteria: it (i) is 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 a reported mRNA half-life, and (v) has adequate MNase-seq coverage.

Genes whose ORFs are less than 500 bp (Supplemental Fig. 13A) long were omitted in order to ensure valid “gene body” calculations between [TSS, +500]. TSS annotations were determined by Park *et al.* (2014). For five genes, *SUL1*, *SUL2*, *MET32*, *HSP26*, and *BDS1*, we manually annotated the TSS to be consistent with the RNA-seq data in this study. We required a half-life for each gene in order to estimate 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 fragment whose center is at this position. MNase coverage was then defined as the number of covered positions in this window divided by the length of the window, 2,000 bp. Genes with MNase coverage below 0.85 ($n=109$) were excluded from further analysis (Supplemental Fig. 13B).

Defining classes of MNase-seq fragments and measures of their occupancy

MNase-seq fragments can be associated with different DNA-binding factors of the basis of their length (Supplemental Fig. 14). To summarize the chromatin occupancy of different factors around genes, fragments were first filtered into two classes: fragments associated with nucleosomes, those between 144–174 bp long, and fragments associated with smaller factors, those less than 100 bp long. In determining these lengths, we made use of two reference data sets, as described next.

Nucleosomal fragment lengths were determined by examining the distribution of MNase-seq fragments prior to cadmium treatment around the top 2,500 unique nucleosome positions reported by a highly sensitive chemical assay (Brogaard *et al.* 2012). In our MNase-seq data, the distribution of fragment lengths at these sites had a clear mode at 159 bp; we chose a ± 15 bp interval around this mode to capture most of the nucleosomal fragments, resulting in the final 144–174 bp range.

As for fragments associated with smaller factors, because prior studies have found clear enrichment of small fragments at Abf1 sites (Henikoff *et al.* 2011), we examined the distribution of our fragments prior to cadmium treatment around 279 Abf1 binding sites, as determined by phylogenetic conservation and motif discovery, obtained from http://fraenkel-nsf.csbi.mit.edu/improved_map/p001_c2.gff (MacIsaac *et al.* 2006). In our MNase-seq data, most of the fragments at these sites were shorter than 100 bp (mode: 75 bp), so those were classified as small fragments.

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

The occupancy of a class of fragments within a particular region is computed simply as the number of fragments of that class whose centers lie within that region.

Computing chromatin scores with cross-correlation kernels

Some chromatin statistics require more spatial precision than occupancy provides, for example when determining a factor’s position or organization. In these cases, we used cross-correlation scores in a similar manner to that described in Tripuraneni *et al.*

(2019). Around each gene's TSS, a per-bp cross-correlation score was computed to smooth positional variation and filter out non-relevant fragments. We constructed three two-dimensional cross-correlation kernels: an idealized, well-positioned nucleosome kernel (Supplemental Fig. 15A), a clearly bound small factor kernel (Supplemental Fig. 15B), and a triple-nucleosome gene body summary kernel (Supplemental Fig. 15C). 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. 15D).

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), as described in the previous subsection.

To summarize the gene body chromatin as a whole, a triple nucleosome kernel was constructed to dampen the effect of the +1 nucleosome becoming more poised to be well-positioned (Mavrich *et al.* 2008; 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. 15E).

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.

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

433 The probability of each outcome is estimated using the triple nucleosome cross-correlation
434 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.

$$\text{cross}_{\text{nuc}}(i) = \text{nucleosome cross-correlation at } i$$

$$\hat{p}(X = x_i) = \frac{\text{cross}_{\text{nuc}}(i)}{\lambda}, \quad \text{where } \lambda = \sum_i^n \text{cross}_{\text{nuc}}(i)$$

Using this random variable, a score was computed for each gene to define its “nucleosome disorganization” using information entropy (Supplemental Fig. 15F):

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

435 *Calling +1, +2, +3 nucleosomes and linking them over time*

436 Nucleosomes were called using peaks of the nucleosome cross-correlation scores local
437 to each gene’s TSS. Peaks within a 1000 bp window around the TSS were sorted by
438 score. The position with the greatest peak score was labeled as a nucleosome center
439 and removed. Positions within 80 bp were also removed. This procedure was repeated
440 until all peak positions were removed and nucleosomes called for this 1000 bp window.

441 “Linked” nucleosomes are defined as nucleosomes across the time course that nom-
442 inally represent the same underlying nucleosome even though its position may have
443 shifted or become more or less fuzzy. Nucleosomes were linked across time points
444 using a nearest-neighbor approach. In a greedy manner, the most well-positioned
445 nucleosome (lowest disorganization score) was considered first. The position of this
446 nucleosome was used to identify the linked nucleosomes in previous and subsequent
447 time points by considering the nearest nucleosome in each of the respective time points
448 within 100 bp of the original nucleosome’s position.

Each gene's +1 nucleosome was called by identifying the linked nucleosome closest to the TSS. The +2 and +3 nucleosomes were computed as the next nucleosomes at least 80 bp downstream from the preceding one.

Analyzing Gene Ontology enrichment

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 corrected using the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995).

Identifying transcription factor binding sites

TF binding sites were called using the small fragment cross-correlation scores in each gene promoter. The cross-correlation scores at each position and time point were sorted by score. The position with the greatest score was removed and labeled as a small fragment peak. Positions within 50 bp of the peak at any time point were also removed. This procedure was repeated until all positions were removed for each gene promoter. A small fragment occupancy value 100 bp around each peak was computed at each time point to identify positions with the greatest change in binding.

The motif finder FIMO (Grant *et al.* 2011) was run against each called peak position against the motif database from MacIsaac *et al.* (2006) using the default p-value threshold. Selected binding sites with supporting literature were annotated on typhoon plots.

Estimating transcription rates

As previously described in Cashikar *et al.* (2005); Rabani *et al.* (2011); Yang *et al.* (2003), transcription rates were computed by incorporating mRNA decay rates into

472 difference equations describing zero-order growth with first-order decay. Details of
473 the procedure used to compute these transcription rates are provided in Supplemental
474 Method S2.

475 *Clustering genes based on chromatin measures*

476 Genes with the greatest increase in average small fragment occupancy or average nucle-
477 osome disorganization were chosen for clustering. The top 500 genes for each measure
478 were combined into a final set of 832 (fewer than 1000 because many genes were in
479 both sets).

480 Clustering was performed in SciPy (Virtanen *et al.* 2020) using hierarchical clus-
481 tering on the basis of pair-wise Euclidean distance between z-normalized measures of
482 change in small fragment occupancy and nucleosome disorganization. Ward linkage
483 was chosen for its efficient approximation to the minimal sum of squares objective
484 (Ward 1963). Eight clusters were ultimately chosen to balance the interpretability of
485 fewer clusters with the significance of identified GO terms in smaller and more homo-
486 geneous but more numerous clusters.

487 *Identifying and quantifying antisense transcripts*

488 TSSs and transcription end sites (TESs) for antisense transcripts were determined using
489 RNA-seq pileup, the number of reads covering a genomic position. To increase signal
490 and decrease noise, at each genomic position we added the antisense pileup values
491 across time points to produce a cumulative pileup, and then smoothed that with a
492 Gaussian kernel.

493 Starting with the highest cumulative pileup value within a gene's transcript bound-
494 ary on the antisense strand, the antisense TSS and TES were identified by progres-

sively searching upstream and downstream, respectively, to identify the positions at which the cumulative pileup values were minimized (Supplemental Fig. 12A). Antisense transcripts were not called if they did not meet a minimum threshold of pileup at any position within the transcript boundary.

For the 667 genes where an antisense transcript could be called (Supplemental Fig. 12B), antisense transcription levels were quantified using a TPM calculation (Wagner *et al.* 2012) for strand-specific RNA-seq reads on the antisense strand within the respective antisense transcript boundaries. We also computed nucleosome disorganization and promoter occupancy chromatin measures relative to these called antisense transcripts, as previously described for the sense strand.

Predicting transcript levels using Gaussian process regression models

Gaussian process regression models were constructed to predict the $\log_2$ transcript level for each time point using the $\log_2$ transcript level and features of the chromatin at 0 minutes, along with 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 nucleosome occupancy within the promoter and within the gene body, small fragment occupancy within the gene body, +1, +2, and +3 nucleosome position shift relative to 0 min (Supplemental Fig. 11), and measures of chromatin relative to called antisense transcripts (Supplemental Fig. 12).

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.

519 The length scale bounds and noise parameters were determined empirically through a
520 sensitivity analysis on a subset of the data.

521 Promoter occupancy and nucleosome disorganization measures were each log trans-
522 formed to yield an approximately normal distribution. Then, each chromatin measure
523 (including nucleosome shift) was z-normalized to allow the RBF length parameter to
524 be successfully approximated.

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

527 **Data Accession**

528 All raw and processed sequencing data generated in this study have been submitted
529 to the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>)
530 under accession number GSE153609.

531 Code to reproduce the results in this study is included in Supplemental Code and
532 available on GitHub (<https://github.com/HarteminkLab/cadmium-paper>).

533 **Acknowledgments**

534 The authors would like to thank Dr. Yulong Li and Dr. Greg E. Crawford for their sug-
535 gestions and helpful comments during the development of this research. The work was
536 supported by the following grants from the NIH National Institute of General Medical
537 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.

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 (red triangle), Met4 (green triangle), and Met32 (obscured by green triangle). **(B)** Depiction of the chromatin dynamics for *HSP26*. (1) Before treatment, nucleosomes are well-positioned. (2) Between 15–30 min, nucleosomes are evicted from the promoter region and replaced by transcription-related proteins and complexes. (3) By 60–120 min, nucleosomes are fuzzy and polymerases are actively transcribing *HSP26*. **(C)** Heatmap of differential cross-correlation values of *HSP26* through the 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).

538

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 small fragment occupancy and $\log_2$ fold-change in transcription rate, each averaged over the time course. **(D)** Scatter plot of relationship between change in nucleosome disorganization and $\log_2$ fold-change in transcription rate, each averaged over the time course.

Figure 4. GO enrichment analysis of genes with highly dynamic chromatin recovers established cadmium response pathways. **(A)** Top 8 categories resulting from GO enrichment analysis of 300 genes with greatest decrease in small fragment occupancy, nucleosome disorganization, and combined chromatin score. Translation-related genes are recovered with significant FDR. **(B)** Top 8 categories resulting from GO enrichment analysis of 300 genes with greatest increase in small fragment occupancy, nucleosome disorganization, and combined chromatin score. Genes involved with stress response, sulfur assimilation, and protein folding pathways are recovered with significant FDR.

Figure 5. Chromatin and transcription dynamics detail Met4 and Met32 functional activation, induction of sulfur genes, and subsequent regulation. **(A)** The Met4 complex activates cascading sulfur pathways required for cadmium chelation and also activates its negative regulator 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 small fragment occupancy (for *MET28*) and nucleosome disorganization (for *MET32*). Sulfur sparing isoforms occur as isoenzyme pairs; members of each pair exhibit inverse chromatin dynamics (most pronounced between *PDC6* and *PDC1*). Nearly all of the sulfur assimilation pathway members show a dramatic increase in small fragment occupancy and nucleosome disorganization. **(C)** Scatter plot of average change in small fragment occupancy and average change in nucleosome disorganization. Chromatin dynamics in sulfur-related genes may manifest primarily in a single measure of the chromatin, as with *MET32* (blue triangle), *MET30* (gray circle), and *PDC6/PDC1* (violet), or in both small fragment 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 gene *MET32*, and sulfur assimilation genes (orange line represents mean and light orange region represents full range of values across all seven genes). The disorganization for Met4 complex cofactor *MET32* 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.

539

Figure 6. Small fragment occupancy in the promoter and gene body nucleosome disorganization reveal stress response pathway timing and patterns with antisense transcription. **(A)** Hierarchical clustering of 832 genes in the union of the 500 with greatest increase in average small fragment occupancy and the 500 with greatest increase in average nucleosome disorganization. Clusters 6–8 contain genes exhibiting anti-correlated chromatin dynamics. **(B)** GO enrichment analysis shows clusters 1 and 2 are enriched for genes in sulfur metabolism and protein refolding pathways, respectively. **(C)** Median (black dot) and interquartile range (red bar) of antisense transcript levels for genes within each cluster across the time course. Cluster 6 genes display a marked increase in antisense transcripts, perhaps explaining why the cluster exhibits increased nucleosome disorganization despite decreased small fragment occupancy in panel 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 throughout the time course. **(C)** Counts of genes that exhibit decreased, unchanged, and increased sense and antisense transcripts from 0–120 min. Genes in each category of sense transcription exhibit positively skewed enrichment of antisense transcripts.

Figure 8. Small fragment occupancy in the promoter reveals transcription factor binding dynamics implicated in cadmium stress response. **(A)** Scatter plot of the 0–60 min occupancy change for 2,119 small fragment peaks identified in gene promoters. 50 peaks increased in occupancy by at least double (red), while 34 peaks decreased by at least half (blue). **(B)** Average change in occupancy for promoter peaks per FIMO-assigned TF. TFs are labeled as increased/decreased (red/blue) if the absolute value of their average log-fold change exceeds 0.1. TFs with the greatest increase in binding occupancy include the iron homeostasis regulators Aft1/Aft2, sulfur pathway regulators Met4/Met32, glycolytic activators Gcr1/Gcr2, and general stress responders Msn2/Msn4. **(C)** Scatter plot of occupancy of Aft1 (turquoise circle) and Aft2 (blue X) at 0 min and 60 min. Aft1 and Aft2 exhibit genome-wide enrichment in binding at 60 min compared to 0 min, particularly in the promoters of a few genes like *SER33*. **(D)** Typhoon plot of *SER33* shows small fragment enrichment at Aft1/Aft2 (red triangle) and Gcr2 (blue triangle, mostly obscured by red triangle) motifs as well as near Met32 (yellow triangle) and Met4 (green triangle) motifs.

540

Figure 9. 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

- Albers E, Laizé V, Blomberg A, Hohmann S, Gustafsson L. 2003. Ser3p (Yer081wp) and Ser33p (Yil074cp) are phosphoglycerate dehydrogenases in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **278**: 10264–10272.
- 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.
- 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.

- 588 Faller P, Kienzler K, Krieger-Liszkay A. 2005. Mechanism of Cd²⁺ toxicity: Cd²⁺
589 inhibits photoactivation of Photosystem II by competitive binding to the essential
590 Ca²⁺ site. *Biochim. Biophys. Acta* **1706**: 158–164.
- 591 Fauchon M, Lagniel G, Aude JC, Lombardia L, Soularue P, Petat C, Marguerie G, Sen-
592 tenac A, Werner M, Labarre J. 2002. Sulfur sparing in the yeast proteome in response
593 to sulfur demand. *Mol. Cell* **9**: 713–723.
- 594 Franzmann TM, Menhorn P, Walter S, Buchner J. 2008. Activation of the chaperone
595 Hsp26 is controlled by the rearrangement of its thermosensor domain. *Mol. Cell* **29**:
596 207–216.
- 597 Geisberg JV, Moqtaderi Z, Fan X, Oszolak F, Struhl K. 2014. Global analysis of mRNA
598 isoform half-lives reveals stabilizing and destabilizing elements in yeast. *Cell* **156**:
599 812–824.
- 600 Grant CE, Bailey TL, Noble WS. 2011. FIMO: Scanning for occurrences of a given motif.
601 *Bioinformatics* **27**: 1017–1018.
- 602 Halimaa P, Blande D, Baltzi E, Aarts MGM, Granlund L, Keinänen M, Kärenlampi SO,
603 Kozhevnikova AD, Peräniemi S, Schat H, Seregin IV, Tuomainen M, Tervahauta AI.
604 2019. Transcriptional effects of cadmium on iron homeostasis differ in calamine
605 accessions of *Noccaea caerulescens*. *Plant J.* **97**: 306–320.
- 606 Hartwig A. 2001. Zinc finger proteins as potential targets for toxic metal ions: Differ-
607 ential effects on structure and function. *Antioxid. Redox Signal.* **3**: 625–634.
- 608 Henikoff JG, Belsky JA, Krassovsky K, MacAlpine DM, Henikoff S. 2011. Epigenome
609 characterization at single base-pair resolution. *Proc. Natl. Acad. Sci. U. S. A.* **108**:
610 18318–18323.

- 611 Heymann P, Ernst JF, Winkelmann G. 2000. A gene of the major facilitator super-
612 family encodes a transporter for enterobactin (Enb1p) in *Saccharomyces cerevisiae*.
613 *Biometals* **13**: 65–72.
- 614 Hosiner D, Gerber S, Lichtenberg-Fraté H, Glaser W, Schüller C, Klipp E. 2014. Impact
615 of acute metal stress in *Saccharomyces cerevisiae*. *PLoS One* **9**: e83330.
- 616 Hu Z, Killion PJ, Iyer VR. 2007. Genetic reconstruction of a functional transcriptional
617 regulatory network. *Nat. Genet.* **39**: 683–687.
- 618 Huisinga KL, Pugh BF. 2004. A genome-wide housekeeping role for TFIID and a highly
619 regulated stress-related role for SAGA in *Saccharomyces cerevisiae*. *Mol. Cell* **13**:
620 573–585.
- 621 Kaiser P, Flick K, Wittenberg C, Reed SI. 2000. Regulation of transcription by ubiqui-
622 tination without proteolysis: Cdc34/SCF(Met30)-mediated inactivation of the tran-
623 scription factor Met4. *Cell* **102**: 303–314.
- 624 Kawahata M, Masaki K, Fujii T, Iefuji H. 2006. Yeast genes involved in response to
625 lactic acid and acetic acid: Acidic conditions caused by the organic acids in *Sac*-
626 *charomyces cerevisiae* cultures induce expression of intracellular metal metabolism
627 genes regulated by Aft1p. *FEMS Yeast Res.* **6**: 924–936.
- 628 Kim TS, Liu CL, Yassour M, Holik J, Friedman N, Buratowski S, Rando OJ. 2010. RNA
629 polymerase mapping during stress responses reveals widespread nonproductive tran-
630 scription in yeast. *Genome Biol.* **11**: R75.
- 631 Klopfenstein DV, Zhang L, Pedersen BS, Ramírez F, Warwick Vesztrocy A, Naldi A,
632 Mungall CJ, Yunes JM, Botvinnik O, Weigel M, Dampier W, Dessimoz C, Flick P,
633 Tang H. 2018. GOATOOLS: A Python library for Gene Ontology analyses. *Sci. Rep.*
634 **8**: 10872.

- 635 Kornienko AE, Guenzl PM, Barlow DP, Pauler FM. 2013. Gene regulation by the act of
636 long non-coding RNA transcription. *BMC Biol.* **11**: 59.
- 637 Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N,
638 Tikuisis AP, Punna T, Peregrín-Alvarez JM, Shales M, Zhang X, Davey M, Robinson
639 MD, Paccanaro A, Bray JE, Sheung A, Beattie B, Richards DP, Canadien V, Lalev
640 A, Mena F, Wong P, Starostine A, Canete MM, Vlasblom J, Wu S, Orsi C, Collins SR,
641 Chandran S, Haw R, Rilstone JJ, Gandi K, Thompson NJ, Musso G, St Onge P, Ghanny
642 S, Lam MHY, Butland G, Altaf-Ul AM, Kanaya S, Shilatifard A, O'Shea E, Weissman
643 JS, Ingles CJ, Hughes TR, Parkinson J, Gerstein M, Wodak SJ, Emili A, Greenblatt JF.
644 2006. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*.
645 *Nature* **440**: 637–643.
- 646 Kubik S, Bruzzone MJ, Albert B, Shore D. 2017. A reply to “MNase-sensitive complexes
647 in yeast: Nucleosomes and non-histone barriers,” by Chereji et al. *Mol. Cell* **65**:
648 578–580.
- 649 Kulaeva OI, Hsieh FK, Studitsky VM. 2010. RNA polymerase complexes cooperate to
650 relieve the nucleosomal barrier and evict histones. *Proc. Natl. Acad. Sci. U. S. A.* **107**:
651 11325–11330.
- 652 Kuras L, Cherest H, Surdin-Kerjan Y, Thomas D. 1996. A heteromeric complex contain-
653 ing the centromere binding factor 1 and two basic leucine zipper factors, Met4 and
654 Met28, mediates the transcription activation of yeast sulfur metabolism. *EMBO J.*
655 **15**: 2519–2529.
- 656 Kuras L, Rouillon A, Lee T, Barbey R, Tyers M, Thomas D. 2002. Dual regulation of
657 the Met4 transcription factor by ubiquitin-dependent degradation and inhibition of
658 promoter recruitment. *Mol. Cell* **10**: 69–80.

- 659 Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient
660 alignment of short DNA sequences to the human genome. *Genome Biol.* **10**: R25.
- 661 Lee CK, Shibata Y, Rao B, Strahl BD, Lieb JD. 2004. Evidence for nucleosome depletion
662 at active regulatory regions genome-wide. *Nat. Genet.* **36**: 900–905.
- 663 Lenstra TL, Benschop JJ, Kim T, Schulze JM, Brabers NACH, Margaritis T, van de Pasch
664 LAL, van Heesch SAAC, Brok MO, Groot Koerkamp MJA, Ko CW, van Leenen D,
665 Sameith K, van Hooff SR, Lijnzaad P, Kemmeren P, Hentrich T, Kobor MS, Bura-
666 towski S, Holstege FCP. 2011. The specificity and topology of chromatin interaction
667 pathways in yeast. *Mol. Cell* **42**: 536–549.
- 668 Lubliner S, Keren L, Segal E. 2013. Sequence features of yeast and human core pro-
669 moters that are predictive of maximal promoter activity. *Nucleic Acids Res.* **41**: 5569–
670 5581.
- 671 MacIsaac KD, Wang T, Gordon DB, Gifford DK, Stormo GD, Fraenkel E. 2006. An im-
672 proved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioin-*
673 *formatics* **7**: 113.
- 674 Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, Schuster SC, Albert I,
675 Pugh BF. 2008. A barrier nucleosome model for statistical positioning of nucleosomes
676 throughout the yeast genome. *Genome Res.* **18**: 1073–1083.
- 677 McIsaac RS, Petti AA, Bussemaker HJ, Botstein D. 2012. Perturbation-based analysis
678 and modeling of combinatorial regulation in the yeast sulfur assimilation pathway.
679 *Mol. Biol. Cell* **23**: 2993–3007.
- 680 Miller C, Schwalb B, Maier K, Schulz D, Dümcke S, Zacher B, Mayer A, Sydow J, Mar-
681 cinowski L, Dölken L, Martin DE, Tresch A, Cramer P. 2011. Dynamic transcriptome
682 analysis measures rates of mRNA synthesis and decay in yeast. *Mol. Syst. Biol.* **7**:
683 458.

- 684 Miller KA, DiDone L, Krysan DJ. 2010. Extracellular secretion of overexpressed
685 glycosylphosphatidylinositol-linked cell wall protein Utr2/Crh2p as a novel protein
686 quality control mechanism in *Saccharomyces cerevisiae*. *Eukaryot. Cell* **9**: 1669–
687 1679.
- 688 Momose Y, Iwahashi H. 2001. Bioassay of cadmium using a DNA microarray: Genome-
689 wide expression patterns of *Saccharomyces cerevisiae* response to cadmium. *Environ.*
690 *Toxicol. Chem.* **20**: 2353–2360.
- 691 Nadal-Ribelles M, Solé C, Xu Z, Steinmetz LM, de Nadal E, Posas F. 2014. Control of
692 Cdc28 CDK1 by a stress-induced lncRNA. *Mol. Cell* **53**: 549–561.
- 693 Nocetti N, Whitehouse I. 2016. Nucleosome repositioning underlies dynamic gene
694 expression. *Genes Dev.* **30**: 660–672.
- 695 Ouni I, Flick K, Kaiser P. 2010. A transcriptional activator is part of an SCF ubiquitin
696 ligase to control degradation of its cofactors. *Mol. Cell* **40**: 954–964.
- 697 Park D, Morris AR, Battenhouse A, Iyer VR. 2014. Simultaneous mapping of tran-
698 script ends at single-nucleotide resolution and identification of widespread promoter-
699 associated non-coding RNA governed by TATA elements. *Nucleic Acids Res.* **42**: 3736–
700 3749.
- 701 Patton EE, Peyraud C, Rouillon A, Surdin-Kerjan Y, Tyers M, Thomas D. 2000.
702 SCF(Met30)-mediated control of the transcriptional activator Met4 is required for
703 the G(1)-S transition. *EMBO J.* **19**: 1613–1624.
- 704 Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Pret-
705 tenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M,
706 Perrot M, Duchesnay E. 2011. Scikit-learn: Machine learning in Python. *Journal of*
707 *Machine Learning Research* **12**: 2825–2830.

- 708 Pereira Y, Lagniel G, Godat E, Baudouin-Cornu P, Junot C, Labarre J. 2008. Chromate
709 causes sulfur starvation in yeast. *Toxicol. Sci.* **106**: 400–412.
- 710 Petti AA, McIsaac RS, Ho-Shing O, Bussemaker HJ, Botstein D. 2012. Combinatorial
711 control of diverse metabolic and physiological functions by transcriptional regulators
712 of the yeast sulfur assimilation pathway. *Mol. Biol. Cell* **23**: 3008–3024.
- 713 Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, Olson S, Weinberg D,
714 Baker KE, Graveley BR, Collier J. 2015. Codon optimality is a major determinant of
715 mRNA stability. *Cell* **160**: 1111–1124.
- 716 Rabani M, Levin JZ, Fan L, Adiconis X, Raychowdhury R, Garber M, Gnirke A, Nusbaum
717 C, Hacohen N, Friedman N, Amit I, Regev A. 2011. Metabolic labeling of RNA un-
718 covers principles of RNA production and degradation dynamics in mammalian cells.
719 *Nat. Biotechnol.* **29**: 436–442.
- 720 Ramachandran S, Ahmad K, Henikoff S. 2017. Transcription and remodeling produce
721 asymmetrically unwrapped nucleosomal intermediates. *Mol. Cell* **68**: 1038–1053.e4.
- 722 Reja R, Vinayachandran V, Ghosh S, Pugh BF. 2015. Molecular mechanisms of riboso-
723 mal protein gene coregulation. *Genes Dev.* **29**: 1942–1954.
- 724 Rhee HS, Pugh BF. 2012. Genome-wide structure and organization of eukaryotic pre-
725 initiation complexes. *Nature* **483**: 295–301.
- 726 Schwabish MA, Struhl K. 2004. Evidence for eviction and rapid deposition of histones
727 upon transcriptional elongation by RNA polymerase II. *Mol. Cell. Biol.* **24**: 10111–
728 10117.
- 729 Shivaswamy S, Iyer VR. 2008. Stress-dependent dynamics of global chromatin remod-
730 eling in yeast: Dual role for SWI/SNF in the heat shock stress response. *Mol. Cell.*
731 *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 reversible by NHEJ-mediated repair in the absence of DNA replication. *bioRxiv* page 866673.

- 755 Vance KW, Ponting CP. 2014. Transcriptional regulatory functions of nuclear long non-
756 coding RNAs. *Trends Genet.* **30**: 348–355.
- 757 Venters BJ, Wachi S, Mavrich TN, Andersen BE, Jena P, Sinnamon AJ, Jain P, Roller NS,
758 Jiang C, Hemeryck-Walsh C, Pugh BF. 2011. A comprehensive genomic binding map
759 of gene and chromatin regulatory proteins in *Saccharomyces*. *Mol. Cell* **41**: 480–492.
- 760 Vinayachandran V, Reja R, Rossi MJ, Park B, Rieber L, Mittal C, Mahony S, Pugh BF.
761 2018. Widespread and precise reprogramming of yeast protein-genome interactions
762 in response to heat shock. *Genome Res.* **28**: 357–366.
- 763 Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski
764 E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Jarrod
765 Millman K, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey C, Polat İ,
766 Feng Y, Moore EW, Vand erPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I,
767 Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P,
768 Contributors S. 2020. SciPy 1.0: Fundamental algorithms for scientific computing in
769 Python. *Nature Methods* **17**: 261–272.
- 770 Wagner GP, Kin K, Lynch VJ. 2012. Measurement of mRNA abundance using RNA-seq
771 data: RPKM measure is inconsistent among samples. *Theory Biosci.* **131**: 281–285.
- 772 Ward JH. 1963. Hierarchical grouping to optimize an objective function. *J. Am. Stat.*
773 *Assoc.* **58**: 236–244.
- 774 Weiner A, Chen HV, Liu CL, Rahat A, Klien A, Soares L, Gudipati M, Pfeffner J, Regev
775 A, Buratowski S, Pleiss JA, Friedman N, Rando OJ. 2012. Systematic dissection of
776 roles for chromatin regulators in a yeast stress response. *PLoS Biol.* **10**: e1001369.
- 777 Weiner A, Hsieh THS, Appleboim A, Chen HV, Rahat A, Amit I, Rando OJ, Friedman N.
778 2015. High-resolution chromatin dynamics during a yeast stress response. *Mol. Cell*
779 **58**: 371–386.

- 780 Wiederhold E, Gandhi T, Permentier HP, Breitling R, Poolman B, Slotboom DJ. 2009.
781 The yeast vacuolar membrane proteome. *Mol. Cell. Proteomics* **8**: 380–392.
- 782 Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers
783 J, Bähler J. 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at
784 single-nucleotide resolution. *Nature* **453**: 1239–1243.
- 785 Yang E, van Nimwegen E, Zavolan M, Rajewsky N, Schroeder M, Magnasco M, Darnell
786 Jr JE. 2003. Decay rates of human mRNAs: Correlation with functional characteris-
787 tics and sequence attributes. *Genome Res.* **13**: 1863–1872.