Molecular Cell, Volume *58*

Supplemental Information

High-Resolution Chromatin Dynamics during a Yeast Stress Response

Assaf Weiner, Tsung-Han S. Hsieh, Alon Appleboim, Hsiuyi V. Chen, Ayelet Rahat, Ido Amit, Oliver J. Rando, and Nir Friedman

SUPPLEMENTAL INFORMATION

Table of Contents

Figure S1

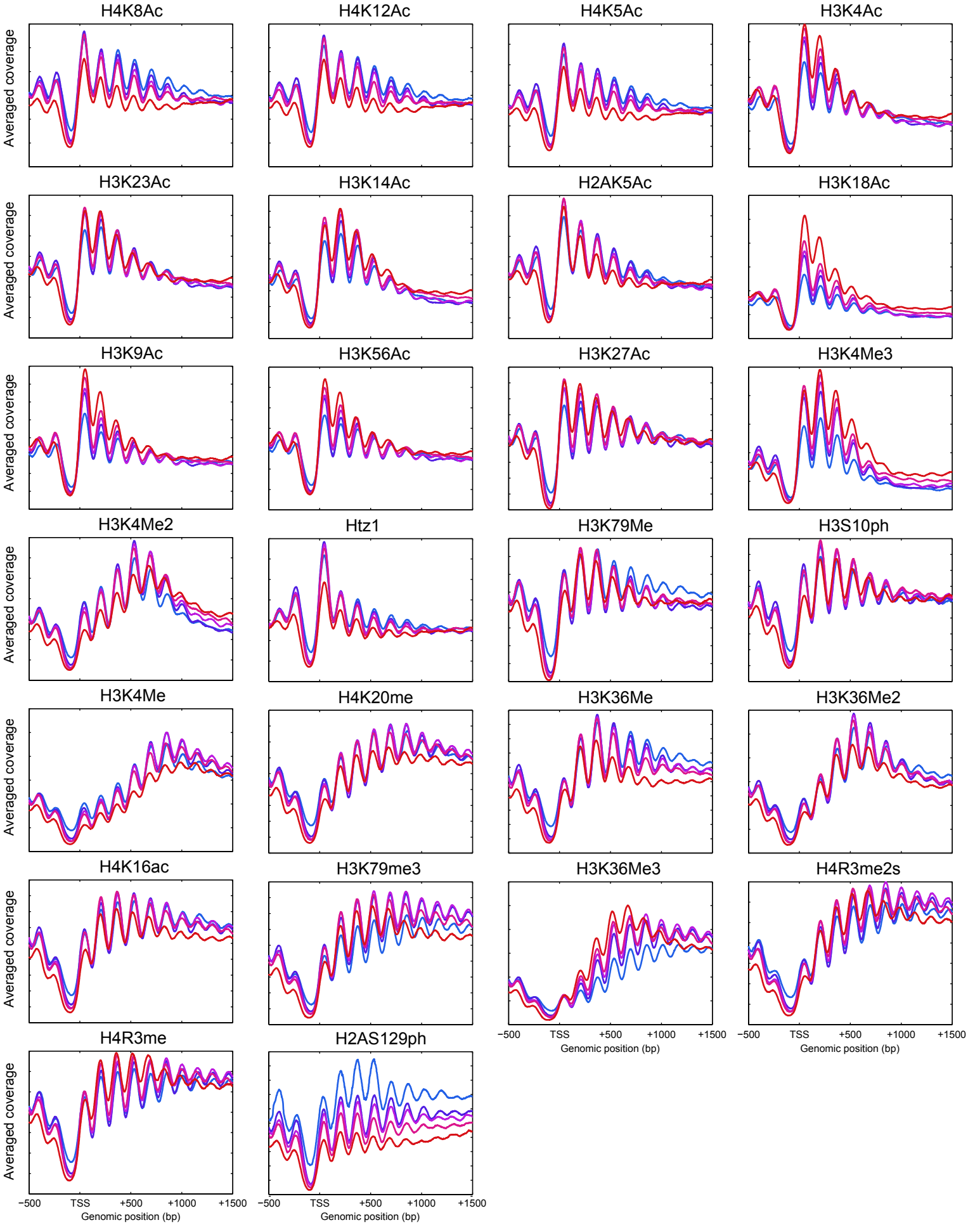
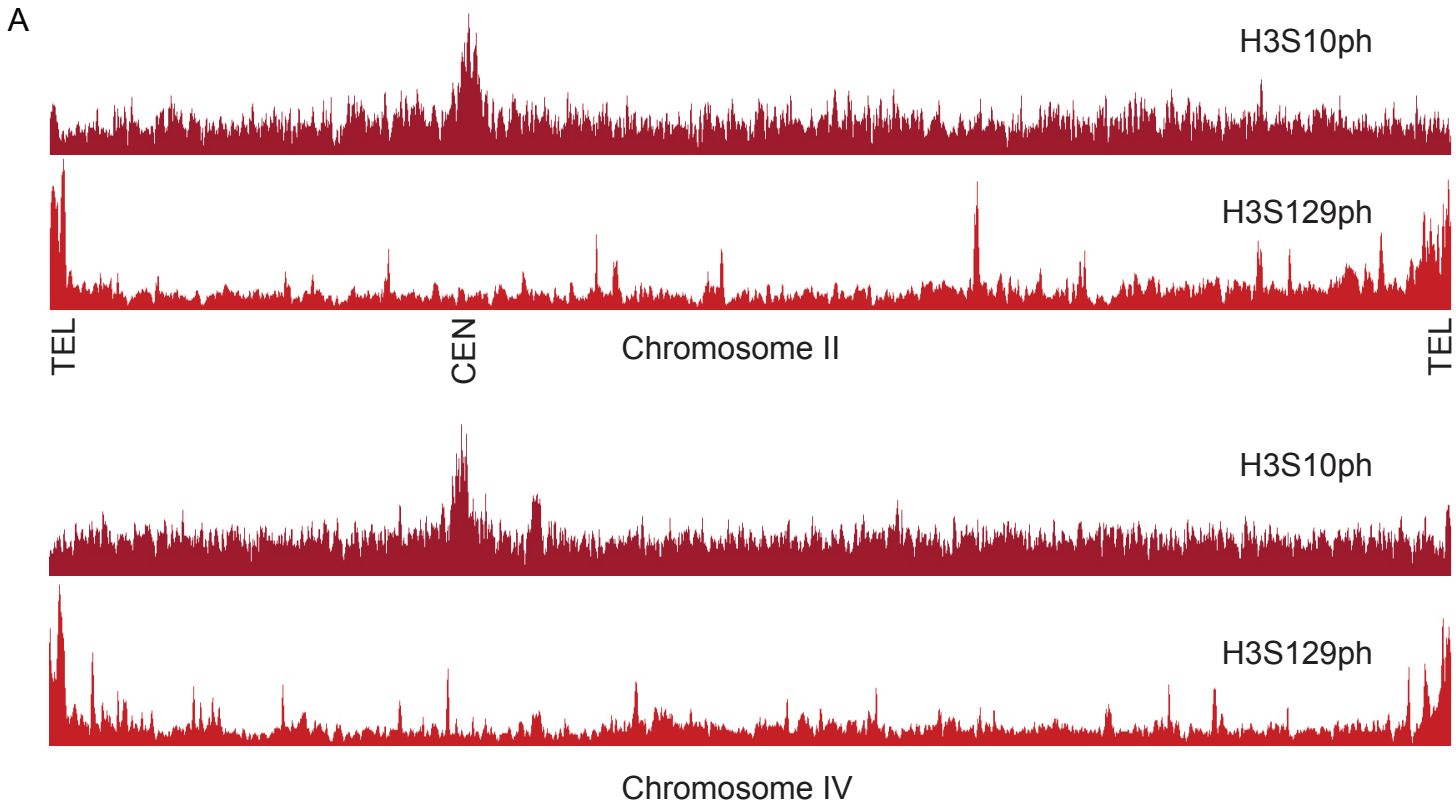

H4K8Ac
H4K12Ac
H4K5Ac
H3K4Ac
H3K23Ac
H3K14Ac
H2AK5Ac
H3K18Ac
H3K9Ac
H3K56Ac
H3K27Ac
H3K4Me3
H3K4Me2
Htz1
H3K79Me
H3S10ph
H3K4Me
H4K20me
H3K36Me
H3K36Me2
H4K16ac
H3K79me3
H3K36Me3
H4R3me2s
H4R3me
H2AS129ph
Averaged coverage
−500
TSS
+500
+1000
+1500
Genomic position (bp)

Figure S1. Metagene profiles at steady-state, Related to Figure 2

For each modification, data were aligned by the transcription start site (TSS) of annotated open reading frames, and grouped according to transcription rate as in **Figures 2E-G**.

Figure S2


A
H3S10ph
H3S129ph
TEL
CEN
Chromosome II
TEL
H3S10ph
H3S129ph
Chromosome IV

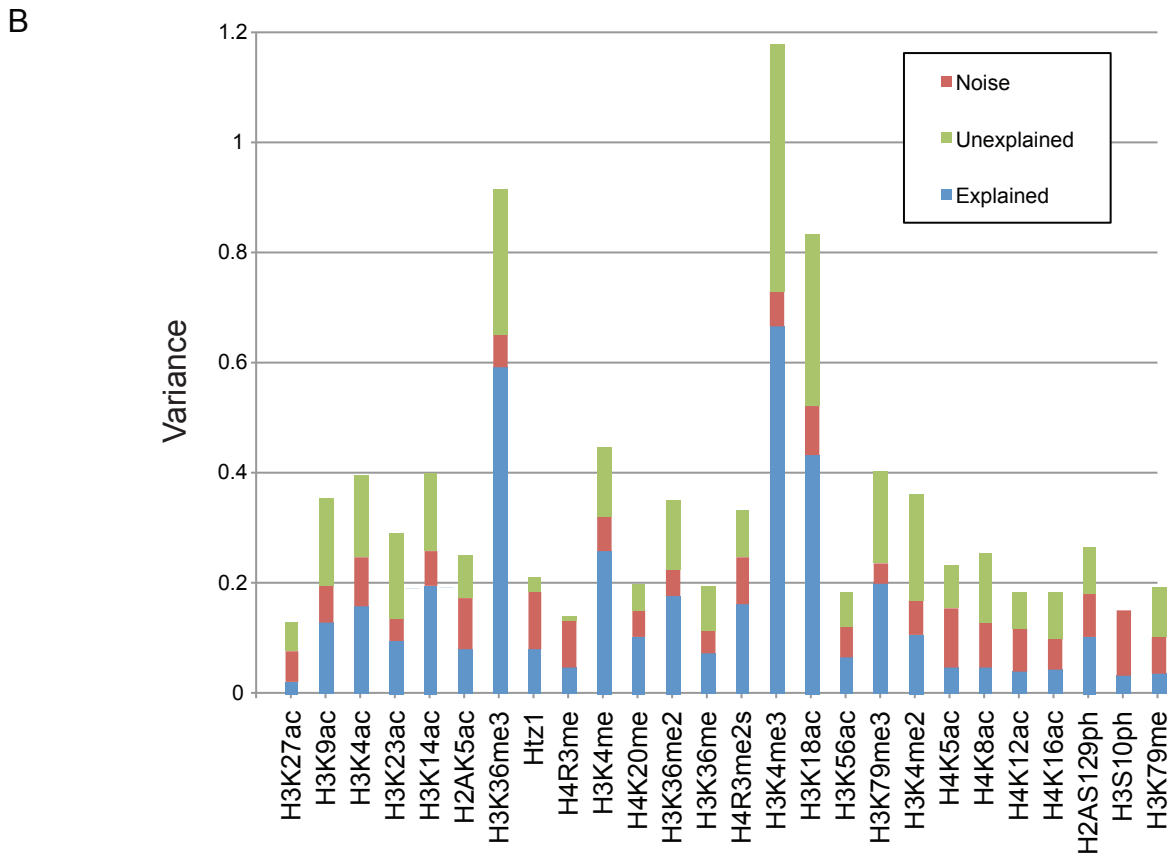

B
Variance
1.2
1
0.8
0.6
0.4
0.2
0
Noise
Unexplained
Explained
H3K27ac
H3K9ac
H3K4ac
H3K23ac
H3K14ac
H2AK5ac
H3K36me3
Htz1
H4R3me
H3K4me
H4K20me
H3K36me2
H3K36me
H4R3me2s
H3K4me3
H3K18ac
H3K56ac
H3K79me3
H3K4me2
H4K5ac
H4K8ac
H4K12ac
H4K16ac
H2AS129ph
H3S10ph
H3K79me

Figure S2. Features contributing to steady-state histone modification patterns, Related to Figure 3

(A) Genome browser views for chromosomes II and IV, showing H3S10ph enriched around centromeres, and H2AS129ph enriched at telomeres, as indicated. **(B)** Contribution of experimental noise to the total variance in different histone modifications. For each modification dataset, we used the difference between adjacent diamide time points to estimate noise in the measurement, assuming that histone marks are relatively stable between adjacent time points. This naturally overestimates the noise in the dataset as yeast are changing their transcriptome and modifying the chromatin template in response to diamide stress. Nonetheless, this provides a reasonable measure of noise given that the features with the lowest noise here are generally associated with transcription-related marks (H3K4me3, etc.) which would be expected based on diamide stress to exhibit the most changes between time points. The higher noise estimated for marks such as H3S10ph and others thus likely reflects true measurement noise, either resulting from relatively nonspecific antibodies, or resulting from widespread marks with lower peak to trough values and thus lower “signal to noise”.

Figure S3

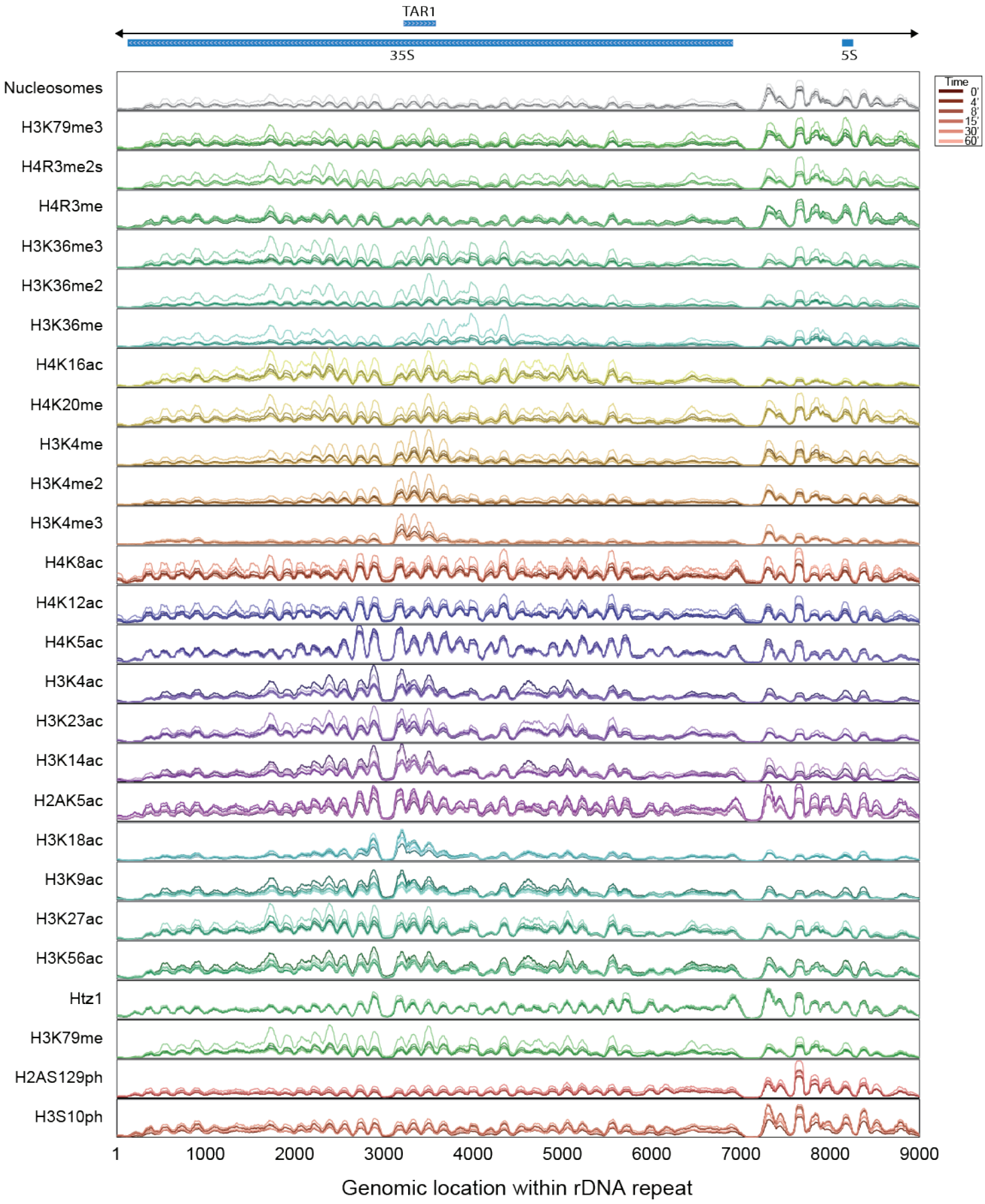
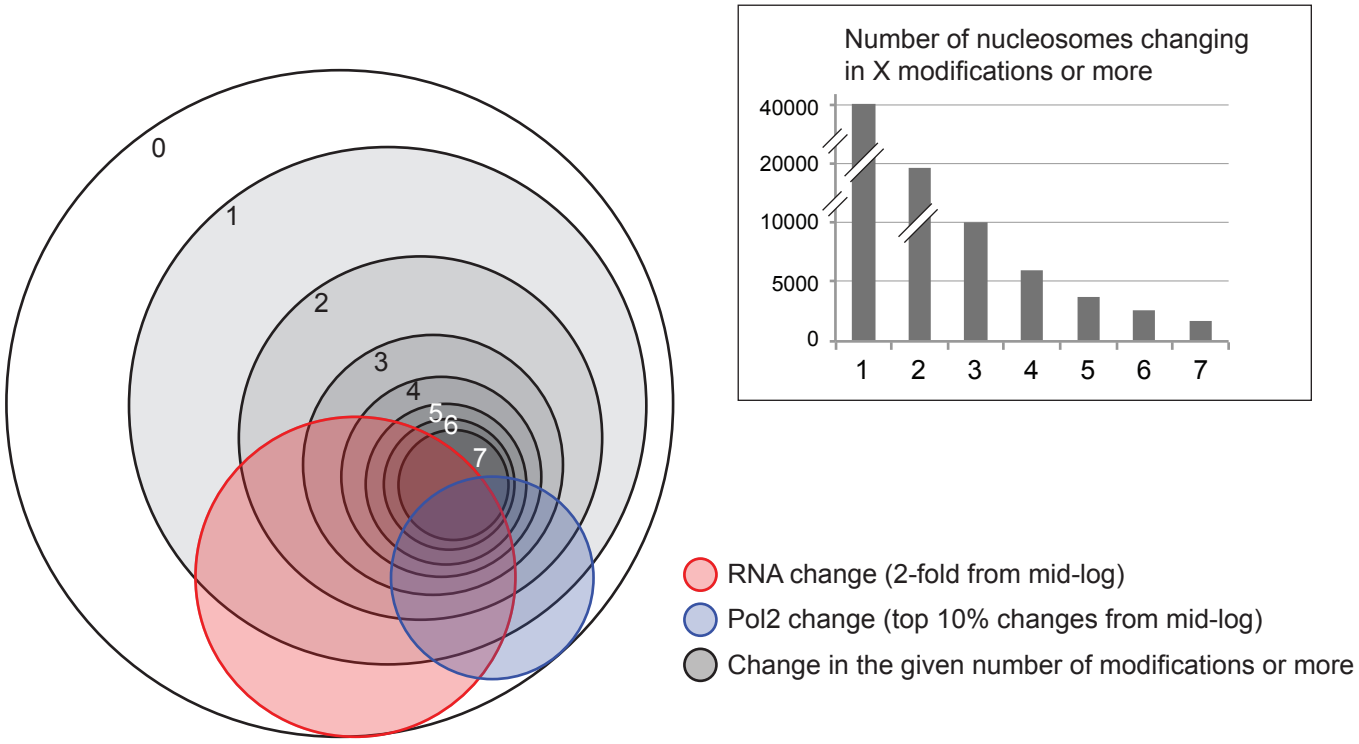

TAR1
35S
5S
Time
0'
4'
8'
15'
30'
60'
Nucleosomes
H3K79me3
H4R3me2s
H4R3me
H3K36me3
H3K36me2
H3K36me
H4K16ac
H4K20me
H3K4me
H3K4me2
H3K4me3
H4K8ac
H4K12ac
H4K5ac
H3K4ac
H3K23ac
H3K14ac
H2AK5ac
H3K18ac
H3K9ac
H3K27ac
H3K56ac
Htz1
H3K79me
H2AS129ph
H3S10ph
1
1000
2000
3000
4000
5000
6000
7000
8000
9000
Genomic location within rDNA repeat

Figure S3. Chromatin landscape of rDNA repeats, Related to Figure 4

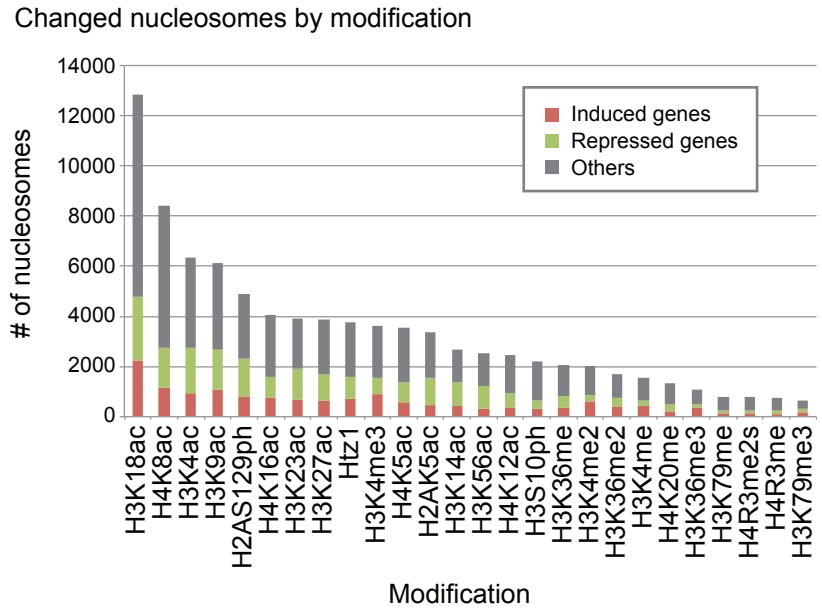
The mapping of nucleosomes and 26 histone modifications in the 9.1kb rDNA repeat region. (top) Browser track showing transcripts from the repeat region. (bottom) Tracks showing enrichment of modifications along the repeat region. Line color saturation in each track ranges from dark (0') to light (60').

Figure S4

A



B



C

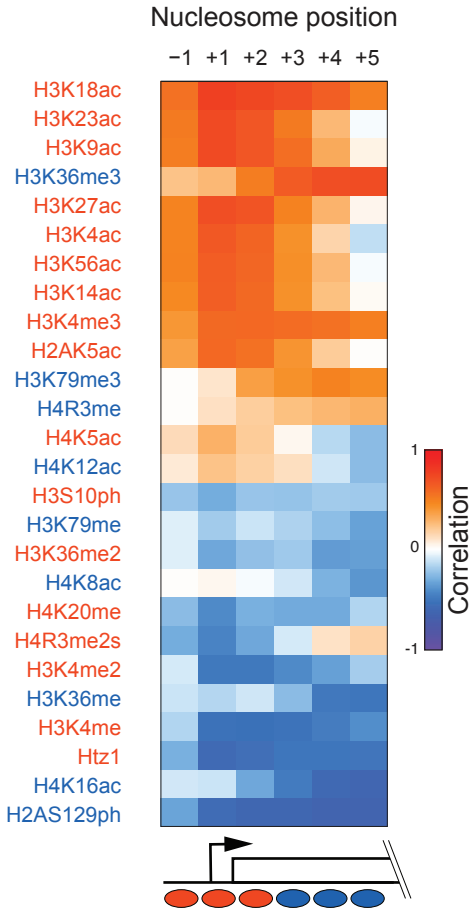


Figure S4. Stress-induced changes in histone modifications, Related to Figure 4

(A) Venn diagram showing the overlap between nucleosomes on genes that changed expression or Pol2 levels (red and blue circle respectively) to nucleosomes that changed in at least 1, 2, …, 7 modifications (gray circles). **Inset**: histogram of the number of nucleosomes in each gray circle. **(B)** The number of nucleosomes that show significant movement in each of the modification. These numbers are broken according to location within induced, repressed genes, and other locations. **(C)** Correlation of change in expression to change of modification (as in Figure 4H) broken by nucleosome position.

Figure S5

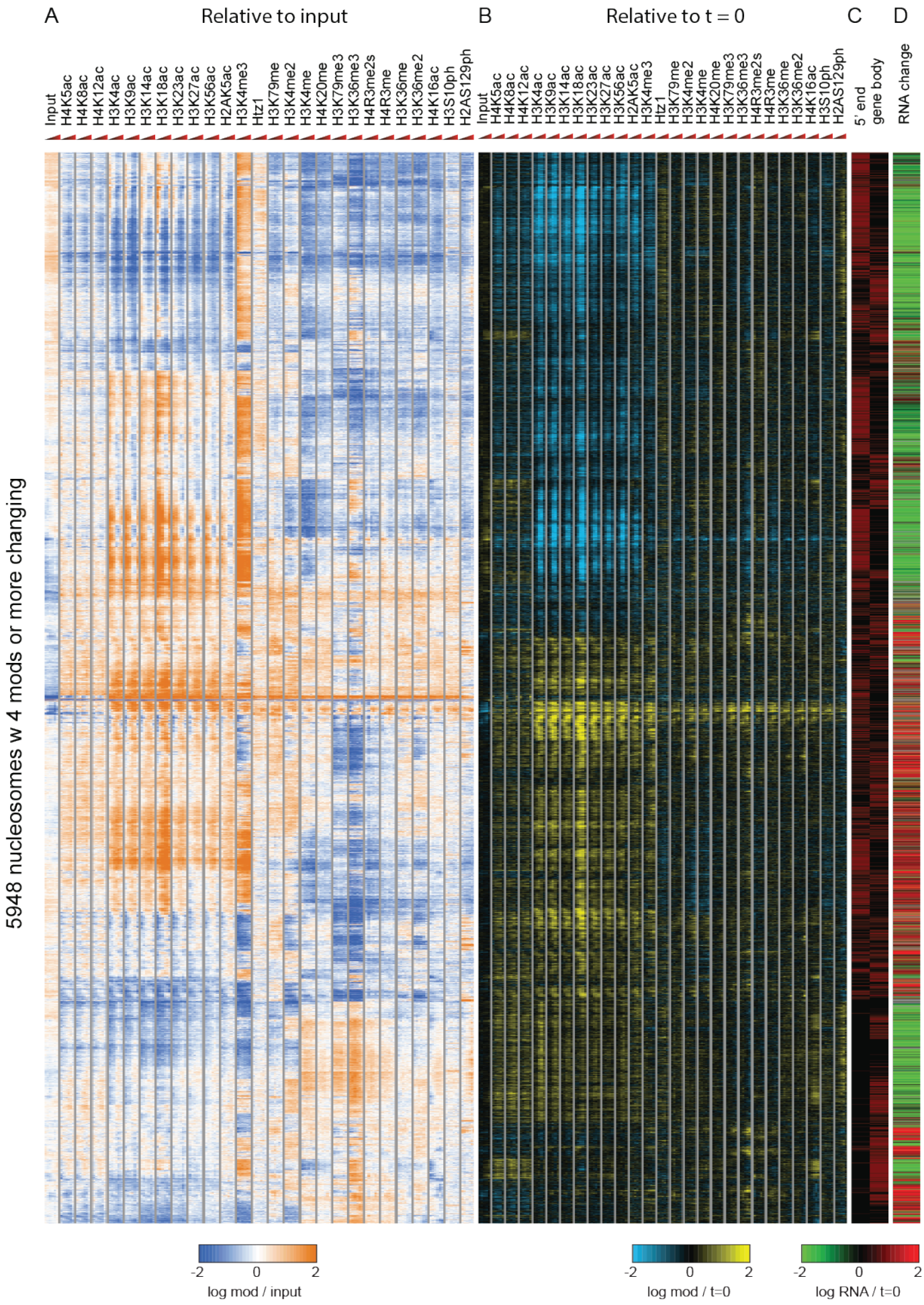

A
Relative to input
B
Relative to t = 0
C
D
Input
H4K5ac
H4K8ac
H4K12ac
H3K4ac
H3K9ac
H3K14ac
H3K18ac
H3K23ac
H3K27ac
H3K56ac
H2AK5ac
H3K4me3
Htz1
H3K79me
H3K4me2
H3K4me
H4K20me
H3K79me3
H3K36me3
H4R3me2s
H4R3me
H3K36me
H3K36me2
H4K16ac
H3S10ph
H2AS129ph
Input
H4K5ac
H4K8ac
H4K12ac
H3K4ac
H3K9ac
H3K14ac
H3K18ac
H3K23ac
H3K27ac
H3K56ac
H2AK5ac
H3K4me3
Htz1
H3K79me
H3K4me2
H3K4me
H4K20me
H3K79me3
H3K36me3
H4R3me2s
H4R3me
H3K36me
H3K36me2
H4K16ac
H3S10ph
H2AS129ph
5' end
gene body
RNA change
5948 nucleosomes w 4 mods or more changing
-2
0
2
log mod / input
-2
0
2
log mod / t=0
-2
0
2
log RNA / t=0

Figure S5. Patterns of stress-induced changes in histone modifications, Related to Figure 4

Hierarchical clustering of histone modification patterns for 5948 nucleosomes with four or more changing marks (see **Supplemental Figure S4A**). Each row is a nucleosome, and columns as follows. **(A)** Time course values of input levels (relative to median levels), and histone marks relative to input. **(B)** Time course values of input and histone marks relative to their levels at t=0. **(C)** Annotation of the nucleosome as 5' or gene body. **(D)** Maximum change in RNA for the associated gene during diamide response. Gray cell denote missing values. Although clustering was performed on the values of (A) and (B), they form coherent clusters in terms of nucleosome position and direction of RNA change.

Figure S6

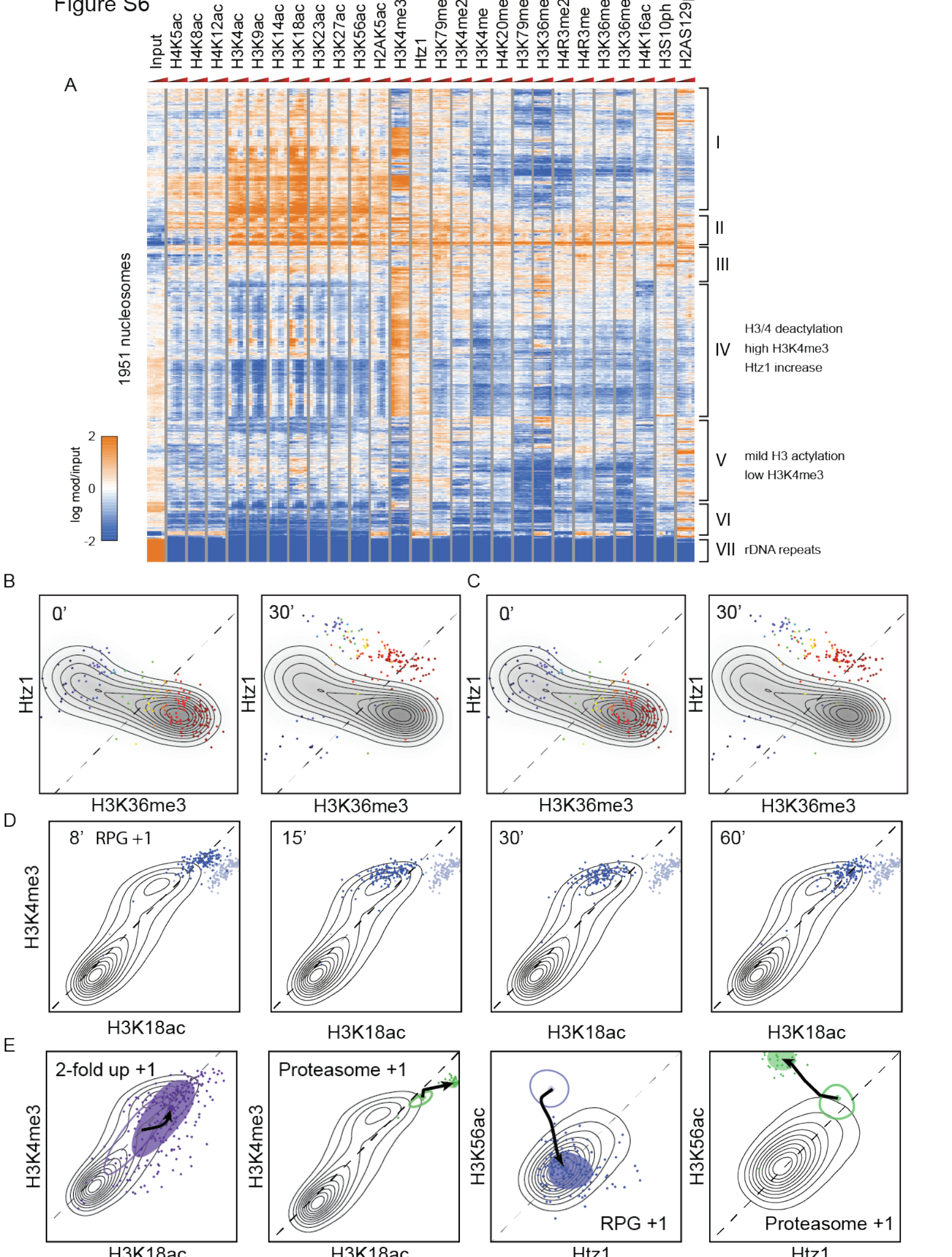

A
Input
H4K5ac
H4K8ac
H4K12ac
H3K4ac
H3K9ac
H3K14ac
H3K18ac
H3K23ac
H3K27ac
H3K56ac
H2AK5ac
H3K4me3
Htz1
H3K79me
H3K4me2
H3K4me
H4K20me
H3K79me
H3K36me
H4R3me2
H4R3me
H3K36me
H3K36me
H4K16ac
H3S10ph
H2AS129
1951 nucleosomes
log mod/input
2
0
-2
I
II
III
IV
H3/4 deactylation
high H3K4me3
Htz1 increase
V
mild H3 actylation
low H3K4me3
VI
VII
rDNA repeats
B
0'
30'
Htz1
H3K36me3
C
0'
30'
Htz1
H3K36me3
D
8' RPG +1
15'
30'
60'
H3K4me3
H3K18ac
E
2-fold up +1
Proteasome +1
RPG +1
Proteasome +1
H3K4me3
H3K56ac
H3K18ac
Htz1

Figure S6. Nucleosomes traversing non-canonical combinations, , Related to Figure 5

(A) Heatmap showing input levels and modification levels (relative to input) for 1915 nucleosomes that leave the high-density region in the 26-dimensional space. Several prominent clusters are noted. **(B-C**). Movements of nucleosomes through 2D modification space. **(B)** The anticorrelation between H3K36me3 and Htz1 (and H2AK5ac) was violated by a number of nucleosomes at late timepoints. **(C)** Nucleosomes that transiently gain the repression-related H2AS129ph despite carrying high levels of H3K36me3 throughout the time course. **(D)** Movement of +1 nucleosomes of Ribosome Protein Genes in the H3K18ac/H3K4me3 space. Blue dots show nucleosomes at the relevant time point, gray dots show nucleosomes at t=0. **(E)** Trajectories for specific sets of nucleosomes are shown, with the t=0 domain being shown as an empty oval, and the stress domain shown as points and a filled oval (as in **Figures 5G and H**).

Figure S7

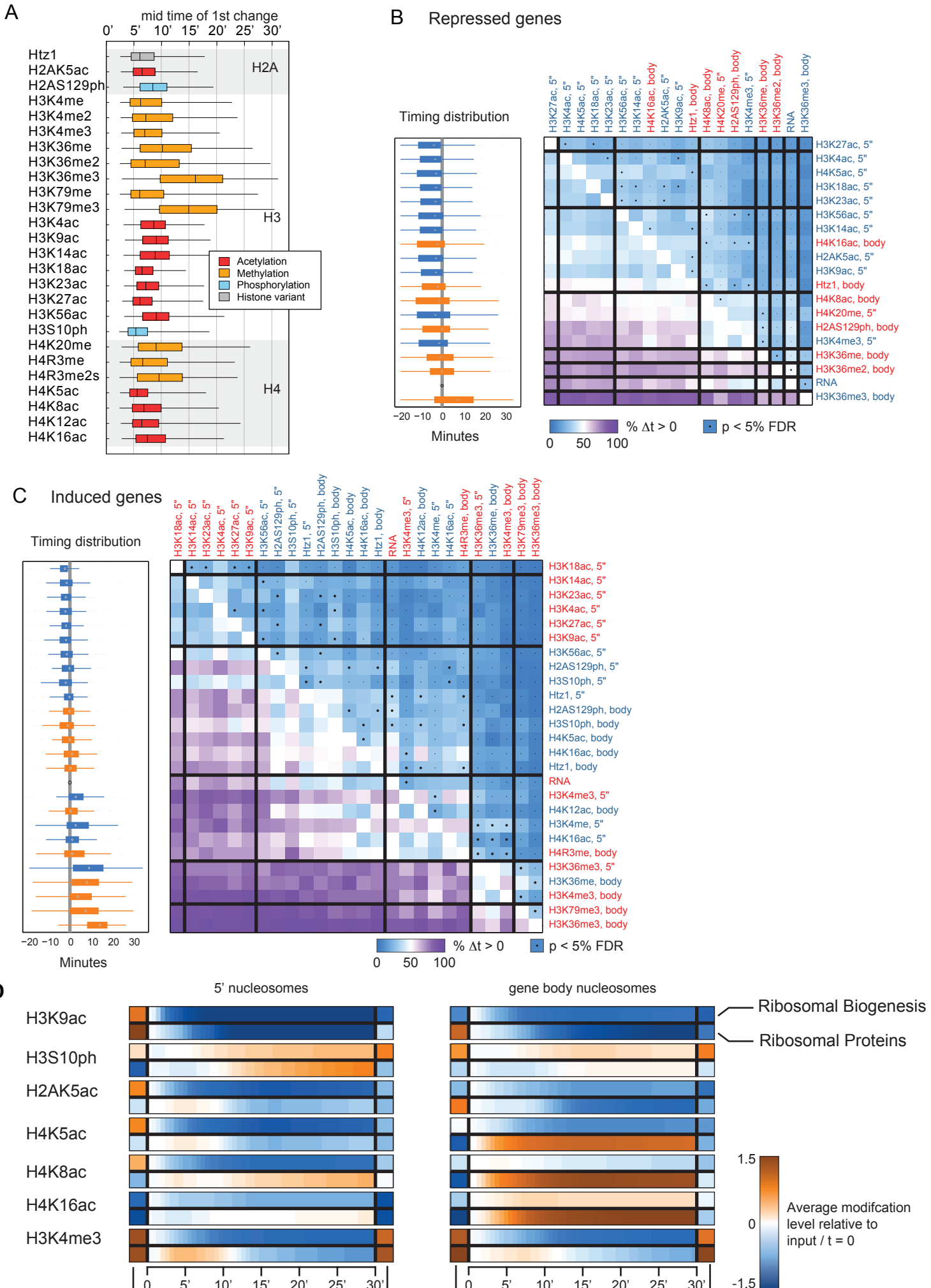

A
mid time of 1st change
0' 5' 10' 15' 20' 25' 30'
Htz1
H2AK5ac
H2AS129ph
H3K4me
H3K4me2
H3K4me3
H3K36me
H3K36me2
H3K36me3
H3K79me
H3K79me3
H3K4ac
H3K9ac
H3K14ac
H3K18ac
H3K23ac
H3K27ac
H3K56ac
H3S10ph
H4K20me
H4R3me
H4R3me2s
H4K5ac
H4K8ac
H4K12ac
H4K16ac
H2A
H3
H4
Acetylation
Methylation
Phosphorylation
Histone variant
B
Repressed genes
Timing distribution
−20 −10 0 10 20 30
Minutes
% Δt > 0
p < 5% FDR
0 50 100
C
Induced genes
Timing distribution
−20 −10 0 10 20 30
Minutes
% Δt > 0
p < 5% FDR
0 50 100
D
5' nucleosomes
gene body nucleosomes
H3K9ac
H3S10ph
H2AK5ac
H4K5ac
H4K8ac
H4K16ac
H3K4me3
Ribosomal Biogenesis
Ribosomal Proteins
0 5' 10' 15' 20' 25' 30'
t=0
t=30
1.5
0
-1.5
Average modifcation level relative to input / t = 0

Figure S7. Timing of changes, Related to Figure 7

(A) t1/2 global by mod Moreover, there are noticeable differences between specific acetylation marks, as H3K18ac, H3K23ac, H3K27ac, and H4K5ac change earlier in the response, while H3K56ac and H4K16ac are slower. Changes in H3S10ph and Htz1 also change as rapidly as the earliest acetylations, while H2AS129ph is relatively slow. **(B)** Repressed genes and **(C)** Induced genes. Right: matrix of relative timing as in **Figure 7D.** Left: box-plot of $t_{1/2}$ for each modification relative to RNA $t_{1/2}$. **(D)** Interpolated time course data for RiBi genes and RPGs for 30 minutes of stress response (as in **Figure 7G**), shown in heat map representation. The left-most and right-most cells denote level relative to genome wide mean at t=0, 30, respectively. The middle row shows changes relative to t=0.

Supplemental Tables

Table S1. Sample details, Related to Figure 1

Contains information for each time series: antibody details (supplier, clone, lot #), experimental batch, sequencing batch, and number of sequenced reads.

Table S2. Nucleosome atlas, Related to Figure 2

Contains information for each nucleosome in the annotated atlas: genomic location of nucleosome center, coverage in the reference mid-log input, and annotation to a position in a gene (if one exists).

Table S3. Normalized modification levels, Related to Figure 2

Table of nucleosome (rows) vs. samples (columns). Each entry is the log2 modification level relative to input. Samples have been normalized as described below.

Table S4. Modification change analysis, Related to Figure 4-5

Table of nucleosomes (rows) vs samples (column). Fore each nucleosome X sample, listed are the $t_{1/2}$, maximum change, interpolation error, and significance analysis for coherent change (see below).

Table S5. Pairwise moving nucleosome analysis, , Related to Figure 5

Number of moving/leaving nucs in each pairwise comparison.

Table S6. Gene sets, Related to Figure 5

List of genes within each gene set in our non-redundant set.

Table S7. Gene set analysis, Related to Figure 5

Gene sets (rows) vs enrichment p-value at different gene positions (see below).

Supplemental Methods

Stress response experiment

All cultures shake in

 # Innova44 shaking incubator

 # 30°C, 220RPM

1. **Culture 400 mL of yeast in 2L baffled flask x 6 (six time points: 0, 4, 8, 15, 30, and 60 min) overnight to OD ~0.55.**

   note 1. Cell number is ~ 0.8 – 1 x 10^7 cells/mL

   note 2. Adjust the total volume of culture based on your experiment design, but the basic principle is to keep the ratio of culture to bottle = 1:5 (400mL : 2000mL)

   note 3. Use a baffled flask to keep constant oxygenation

   note 4. Although the diamide treatment will slow yeast growth, they still grow ~1.2-1.3x at 60-mins point in our system. In order to get constant condition between time points, please check the growth curve of your strain beforehand

2. **Add reagents as following table. Once you add diamide into '60mins-culture', start the timer, and then add diamide into next bottle every 1 min after (ex. Add to 30min-culture at 1min on timer, 15min-culture at 2min on timer, and so on).**

	1M Diamide	**37% Formaldehyde** *30C for 15min*	**2.5M Glycine** *RT for 5min → on ice*
Final conc	*1.5 mM*	*1%*	*125 mM*
Add	*600 uL*	*10.7 mL*	*20.5 mL*
0 min	x	6	21
4 min	4	8	23
8 min	3	11	26
15 min	2	17	32
30 min	1	31	46
60 min	Start → 0 (min)	60	End → 75 (min)

note 1. Freshly prepare 1M diamide stock [MW = 172.19] in TE buffer. If not, store in -20 freezer, don't freeze and thaw over 3 times.

3. **Spin for 5mins @ 4000rpm, 4°C**
4. **Wash cell pellets by 50mL water**
5. **Spin for 5mins @ 4000rpm, 4°C**

MNase-ChIP Protocol (beads-beating)

Solution:

- 0.5mm diameter ZIRCONIA/SILICA beads Cat.11079105z, BioSpec
- 2mL Screw-cap tubes
- Cell breaking buffer(0.1M Tris, pH7.9, 20% glycerol), 4°C
- Sigma protease inhibitor cocktail for fungi (PIC), 100X
- NP Buffer: 0.5 mM spermidine, 1 mM β-ME, 0.075% NP-40, 50 mM NaCl, 10 mM Tris pH 7.4, 5 mM $MgCl_2$, 1 mM $CaCl_2$. **Do NOT include Sorbitol!**
  Ex: 5 ml of NP Buffer:
  10 ul 250 mM spermidine
  3.5 ul of 1:10 (diluted in water) b-ME
  37.5 ul 10% NP-40
  Bring up to 5 ml with MNP buffer.
- Buffer L: 50 mM Hepes-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate.
- Buffer W1: Buffer L with 500 mM NaCl
- Buffer W2: 10 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1mM EDTA
- Buffer Z: 1 M sorbitol, 50 mM Tris pH 7.4
  Ex: 1 L of buffer Z:
  500 ml 2 M sorbitol
  50 ml 1 M Tris-HCl pH 7.4
  450 ml ddH_2O
- TE: 10 mM Tris-HCl pH 8.0, 1 mM EDTA
- "2X" Proteinase K solution: TE with 0.8 mg/ml glycogen, 2 mg/ml proteinase K
- Elution buffer: TE pH 8.0 with 1% SDS, 150 mM NaCl, and 5 mM DTT **Do not add the DTT until just before use**
- Zymolyase solution (10 mg/ml in Buffer Z; lasts up to 2 weeks at 4°C)
- Micrococcal Nuclease (Worthington Biochem): resuspended from lyophilized powder at 20 U/ul in Tris pH 7.4. Aliquot into tubes upon first use and freeze at –80°C.

Protocol:

1. Pour ~1000uL of 0.5mm beads into screw-cap tubes and leave on ice.
2. For each pellet collected from 400mL culture, resuspend with 1000uL of cell breaking buffer (+1X PIC), and then aliquot to 2mL screw-caps tube x2 from step 1.
3. Put 12 tubes in pre-chilled (-20°C) magnetic bead-beating rack and bead-beating for 3 mins twice by Biospec beads beater.
4. Stab a hole at bottom of tube by a heating needle. Move stabbed tube into 5mL big tube and collect sample by spin for 1min @ 700g, 4°C.
5. To collect nuclei, pipette sample to 1.5mL canonical tube and spin for 10mins @ Max speed, 4°C.
6. Get rid of the supernatant. Resuspend the pellet with 2400uL NP buffer.

7. Add appropriate volume of MNase (Note: Titrate it beforehand. Usually getting 90% monomer + 10% dimer + little trimer band would be ok). Briefly mix and incubate for 20mins @ 37°C.
8. Add 24uL of 0.5M EDTA and incubate for 10min @ 65°C to inactivate enzyme activity.
9. Prepare Protein A beads. Pipette 1500ul slurry (250uL per sample) → Spin down for 30sec @ 3000g, 4°C → Wash beads twice in Buffer L → Recover to original volume in Buffer L.
10. Pool the digested material and add the following to the digestion products to simulate Buffer L conditions. **Add the salts before the detergents!** Amounts below are per a **600 ul digestion aliquot; scale accordingly (here is 2400uL, so x4 for each one)**.

Volume (ul)	Component
80	0.5 M HEPES-KOH pH 7.5
22.4	5 M NaCl
6.4	12.5% sodium deoxycholate
80	10% Triton X-100
8	Sigma protease inhibitors

11. Set aside at least 200 ul (~5%) of pool as non-IP control; needed also for gel verification of the MNase digest.
12. Pipette 250uL protein A to each sample and rotate at 4°C on tube rotisserie for one hour.
13. Spin for 30 seconds @ 3000g, 4°C, and transfer 700-800uL of supernatant to each tube containing the appropriate amount of antibody (total is 6 time points x 4 sets = 24 tubes).
14. Incubate with rotation at 4°C for 4 hours to overnight (up to 16 hours).
15. Add 100uL of protein A bead to each tube.
16. Incubate with rotation at 4°C for 1 hour (longer is allowable but not necessary).
17. Spin for 30 seconds at 3000g at 4°C and for subsequent pelleting steps in washes.
18. Wash beads successively with 1 ml of the following buffers, for 5 minutes (on rotisserie at 4°C) each, in the following order: Buffer L (twice), W1 (twice), W2 (twice), and TE pH 8.0 (twice).
19. Incubate and mix the beads in 125ul of elution buffer at 65°C for 10 minutes on thermal mixer. Be sure to add the DTT in the elution buffer (to 5 mM final concentration) beforehand.
20. Pellet the beads by centrifugation at 10000g for 2 minutes and **keep the supernatant**.
21. Repeat Steps 19 and 20 and discard the protein A beads when done.

22. Reverse cross-linking by Proteinase K
23. Extract DNA with PCI
24. Ethanol precipitation of DNA
25. RNAase digestion
26. CIP
27. Clean-up by Minelute

Data Quality and Antibody Quality Control

Although the dataset analyzed here consists of single replicate data, numerous preliminary datasets were gathered for several modifications, including several ChIP-Seq datasets for H3K36me3 as well as several published (Weiner et al., 2012) and unpublished microarray analyses of over ten distinct histone marks during a diamide stress response. All such replicates were highly concordant with the data presented here.

In selecting antibodies for this studies, we analyzed reports of antibody quality (Egelhofer et al., 2011) (http://compbio.med.harvard.edu/antibodies/), and did not consider antibody lots with previous reports of cross-reactivity. For six modifications, we performed ChIP-chip [as detailed in (Weiner et al., 2012)] in strains lacking the modified residue (Dai et al., 2008) and removed those that showed cross-reactivity – all antibodies to H4K91ac were eliminated based on this step, as they gave similar ChIP-chip profiles in wild-type and H4K91A mutants. Finally, we assessed the quality of antibodies which passed these filters by evaluating how different they are from the input distribution. Specifically, we estimated the percent of their variance explained by normalizing them to the input, and conditioning on position and expression information (**Supplementary Tables S2 and S3**). A visual inspection revealed that indeed the ten least explained antibodies are almost identical to the input, with the exception of H3S10ph, which is localized to the centromeres. To quantify the visual inspection, a genome wide correlation score to the input revealed a clear separation between these ten antibodies, which were discarded, and all other antibodies (**Supplementary Table S1**).

Data Processing

The normalization was done in few steps:

1. We generated coverage plots from uniquely mapped reads. Since the data was from mono-nucleosomal fragements, each read was extended to 100bp before computing coverage. The coverage was normalized to 10^7 total number of reads in each sample and smoothed running window averages of 15bp width.
2. To call nucleosome peaks, we first identified local maxima in coverage in each input experiments (all time points X all batches). We then applied a greedy procedure to select the ones with the highest coverage as the centers of nucleosomes, where each selected

peak removes from consideration all other peaks whose center is within +/- 100bp. The selected peaks formed the nucleosome atlas of **Supplementary Table S2.**

3. Coverage estimation. For each nucleosome, we computed the coverage in each experiment as the maximum number of reads covering a single base within 100 bp of the nucleosome center. To prevent overflow in normalization to input, we add 25 "pseudo reads" to all coverage values (in both input and ChIP samples).
4. Each sample was then represented as $\log_2$ of the ratio to input. To take into account batch differences, this ratio was computed as the average of the input signal from the matching batch and the mean of all other inputs.
5. We then applied QQ normalization to each time series (each antibody). This normalization assumes that the distribution of values in in each antibody through time remains the same. The normalization matches quantiles in each time-course to each other (using MATLAB's quantilenorm, version R2013a).
6. The normalized log ratios are reported in **Supplemental Table S3** and used throughout the analysis.
7. Final output of these steps was visually compared to raw genome browser tracks.

TSS mapping and Expression Data

Expression data during diamide stress response was taken from (Gasch et al., 2000). For a subset of analyses, Pol2 ChIP-chip from (Kim et al., 2010) was used instead place of mRNA abundance, with all key conclusions being qualitatively identical using either dataset.

We annotated nucleosome positions along the gene (-1,+1,+2,…,+N) based on TSS mapping data, generated as in (Ni et al., 2010). Briefly, Polyadenylated RNA was treated with bacterial alkaline phosphatase (TAKARA), then decapped using Tobacco Acic Pyrophosphatase (Epicentre). An oligo containing an MmeI site was ligated selectively to previously capped 5' ends of RNA using RNA ligase. After reverse transcription and low cycle amplification using biotinylated primers, MmeI was used to digest 20bp downstream from where the 5' cap had been. This DNA was then isolated with streptavidin beads and ligated to a modified Illumina adaptor. After elution from the beads, TSS sequences were amplified by PCR, cloned, and deep sequenced.

Noise estimation

To estimate the technical noise levels of each ChIP experiment we treated time-point measurements as biological replicate by selecting the time point with the smallest differences to time t=0 as the second replicate for noise estimation.

Regression and sparse regression

We used multiple linear regression analysis to reconstruct histone modifications levels from a collection of features. Feature of the regression are:

(a) Nucleosome position relative to transcription start site (**Supplemental Table S2**).
(b) Mid-log occupancy level taken from the merged MNase input signal (**Supplemental Table S2**).
(c) RNA polymerase levels from published NET-seq data (Churchman and Weissman, 2011). For each nucleosome we counted the number of sense and antisense (AS) NET-seq reads up-to 100 bp from its dyad. Sense/AS were determined based on SGD genes annotations.
(d) Turnover data was taken from (Dion et al., 2007), for each nucleosome we considered the average value from microarray probes with distance of 100 bp for its center. 34830 nucleosomes had at least one probe, the rest were discarded from the analysis.
(e) Positions relative to nearest centromere/telomere in base pairs (log).
(f) Replication Timing was based on (Raghuraman et al., 2001), we assigned timing value in minutes for each nucleosome using linear interpolation of the reported data.

In total, we assigned 6 features to each nucleosome plus it's position along the genome. Finally, we learned the multiple regression coefficients for each genomic position separately.

PCA

PCA analysis was performed using MATLAB's pca method where all 6 time-points were merged to one large matrix (66360 X 6): 398160 X 26.

Detecting nucleosomes at low density regions

To investigate the 26-dimensional modification space, we employed a semi-parametric technique of kernel density estimation. We define the density function at point $x \in R^D$ to be:

$$d(x,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{k \in NN_0} e^{-\sum_j \left(\frac{x_j - x_{kj}}{\sigma}\right)^2}$$

Where σ is the bandwidth of the kernel, NN_0 are the 20 nearest neighbours to x, at t=0, omitting the single nearest neighbor (for stabilization considerations).

To find the optimal bandwidth, we use a cross-validation approach; we randomly draw half of the nucleosomes to form a training set, and estimate the likelihood of the other, unseen, part of the data. The optimal bandwidth is the one that maximizes the likelihood:

$$\sigma^* = \underset{\sigma>0}{\operatorname{argmax}} \prod_{i \notin train} d(x_i, \sigma)$$

Given the optimal bandwidth, σ^*, we can continue to estimate the density with respect to mid-log modification space at all time points.

Compendium of gene sets

We assembled a compendium (**Supplemental Table S6**) of gene sets of functional groups (Ashburner et al., 2000; Dutkowski et al., 2014; Segal et al., 2003), DNA binding data (Harbison et al., 2004; Rhee and Pugh, 2012; Venters et al., 2011), genetic perturbations (Chua et al., 2006; Lenstra et al., 2011; Mnaimneh et al., 2004), and RNA binding data (Gerber et al., 2004). We removed redundant gene sets by selecting a smaller set of representative gene sets such that all gene sets have a Jaccard distance of 0.2 or lower to one of these representatives (keeping ~60% of ~13000 original gene sets).

Gene-sets with rare modification states

For each pairwise modification space, gene-set and nucleosome position we tested whether nucleosomes at the position in genes in this set is over-represented in the low-density region of the pairwise combination (hyper-geometric p-value). We corrected for multiple testing with 5%-FDR, removing non-significant results. We then average the log of these p-values over all the 2D spaces to assign aggregated p-values for each gene-set (**Supplemental Table S6**).

Fit and t1/2 estimation

We have no specific prior on a functional form for the modification responses so we use a non-parametric approach for our estimation. We do assume that the responses are smooth, and that modification levels are at mid-log steady state when t=0.

To estimate the response we introduce steady-state pseudo-measurements at -60 and 120 minutes (values as in t=0) based on previous observations in the literature that the yeast return to baseline transcriptional state after 90min (Gasch et al., 2000; Kim et al., 2010). We then iterate over internal points (4,8,15,30 minutes), leave each one out, and calculate the cubic interpolation (MATLAB's **interp1,** version R2013a). The estimated response is the mean of all these leave-one-out (LOO) interpolated responses.

Given the estimate, the peak change is defined as the point in time which has the maximal absolute change, relative to t=0. After evaluating the peak change, one can define $t_{1/2}$, as the time at which the estimated response reaches half the peak change, and $t^r_{1/2}$, as the time at which the estimated response is at half the return to the value at t=0 ($t^r_{1/2}$ is not necessarily applicable). These time points re estimated using a cubic interpolation resolution of less than .6 minutes, along with a linear interpolation between $t_{1/2}$ flanking time-points.

The error of the estimate is calculated as the mean difference between each of the LOO-interpolated responses and the data point omitted, divided by the standard deviation of the data series. Formally:

$$err(x_i) = \frac{1}{STD(x_i)} \sqrt{\sum_{t\in[4,8,15,30]} \left(y_i^{-t}(t) - x_{it}\right)^2}$$

Where $y_i^{-t}(t)$ is the interpolated response obtained when time point t is omitted, evaluated at time point t. x_i is the measured log-fold change vector (from t=0) of a nucleosome indexed by i, at a certain modification.

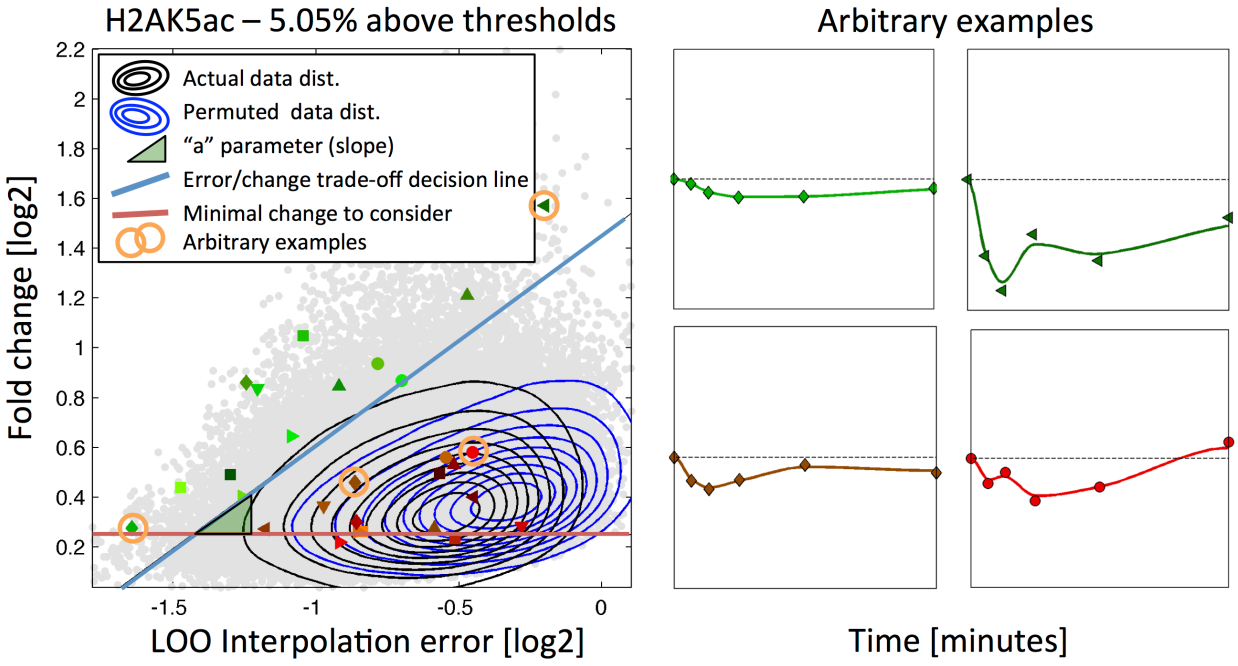
Coherent Movement determination

For each modification, we use a permutation test to estimate which nucleosomes exhibit a coherent change in their response. The null hypothesis is that changes are random fluctuations in measured data, is obtained by permuting the nucleosomal measurements independently at each time point. This maintains the overall distribution of change values but eliminates any

connection between these measurements through time. On this permuted dataset we employ the same procedure described above (LOO interpolation) to obtain error estimates, as described above. Finally, for each nucleosome, i, in each modification, m, we define the following statistic:

$$s_m^i = PC(x_m^i) - a_m \log\left(err(x_m^i)\right)$$

Where $PC(x_m^i)$ is the (log) peak change of nucleosomes i in modification m, $err(x_m^i)$ is the error of the LOO interpolation, and a_m is a global modification constant weighing the relative weight of the two numbers:



Now, for each modification, for a given FDR α, we set the threshold over this s statistic to be the maximal such that at most α% of nucleosomes above the threshold are not from the randomly permuted data set. For most analyses we use an FDR of 10%, a visual inspection reveals that this is generally a stringent cut-off. For the timing-of-events analysis we use a more lenient FDR of 25%, to allow for greater statistical power in downstream analysis, assuming that the selection of specific nucleosomes of specific genes is independent of false discoveries in this s-statistic. To set the a_m per modification, we optimize over the number of nucleosomes passing a certain FDR (10%). While this might increase the actual FDR, the stability of this global constant across modifications and FDR thresholds, and a visual inspection of results, suggests that the optimal constants represent an actual tradeoff between these quantities, rather than an arbitrary number.

Event Pair Statistics

Given the collection of coherent events in a gene set of interest, we next ask "How do these events relate to each other?" We define the precedence of one event, A, over the other, B, with respect to a gene-set G:

$$P_G(A,B) = \frac{1}{|G|}\sum_{g\in G} 1_{\left\{t_{1/2}(A_g) > t_{1/2}(B_g)+\tau\right\}}$$

Or in words – we simply count the fraction of genes in which event A is preceded by event B, plus some confidence interval ($\tau = 1[min]$, this means that any events with a timing difference of less than one minute is ambiguous and ignored).

Also, for each such coherent event pair, a one sided t-test was performed to exclude the possibility of a random timing difference between these events. The t-test associated p-value was collected, and only those p-values that passed an FDR threshold of 5% were further considered. These p-values allow us to define the timing of events (TOE) graph w.r.t to a gene-set G - $\langle V_G, E_G \rangle$, as follows:

$$V_G = coherent\ events\ in\ G$$

$$E_G = \{(A,B) \mid A\ precedes\ B\ significantly\ in\ G\}$$

Order of Events

Note that the TOE graph is a directed acyclic graph (DAG) by definition (and the linearity of time), so it provides us with a partial order of events. Ideally, we would like to find the optimal order of events, π^*, which is consistent with the TOE graph, T, such that:

$$\pi^* = \underset{\pi\in S_{|T|}\subseteq T}{\operatorname{argmax}} \sum_i \sum_{j>i} P(\pi_i, \pi_j)$$

Where $S_{|T|}$ is the permutation set over the vertices of the graph, i.e. over events, and the notation $S_{|T|} \subseteq T$ means that they are consistent with the order dictated by the edges of T. In other words, we are looking for the order of events that maximizes the "overall precedence" of the data, while conforming to precedence significance in the data. Unfortunately, this problem (optimal order w.r.t. an edge weight function, even without constraints) has been shown to be NP-complete problem usually called the LINEAR-ORDER-PROBLEM (or LOP).

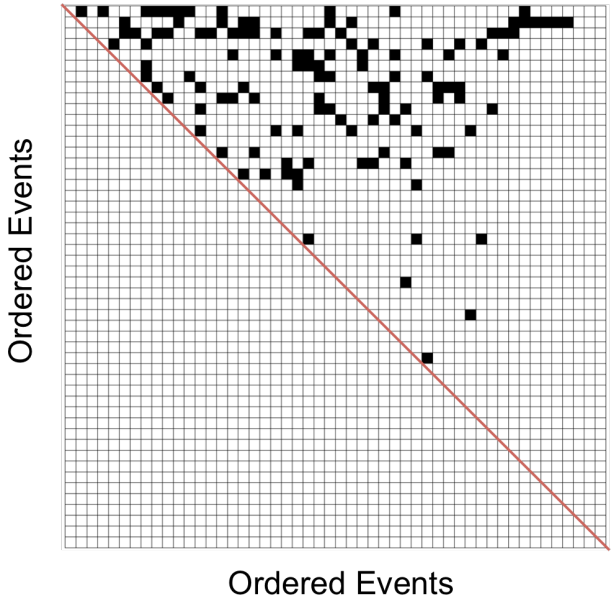
As a heuristic we perform a topological sort of the data (which guarantees that the graph constraints are satisfied), and in the cases where two events are incomparable in the partial

order, we use the overall precedence of these events (w.r.t to all other events) to obtain a total linear order.

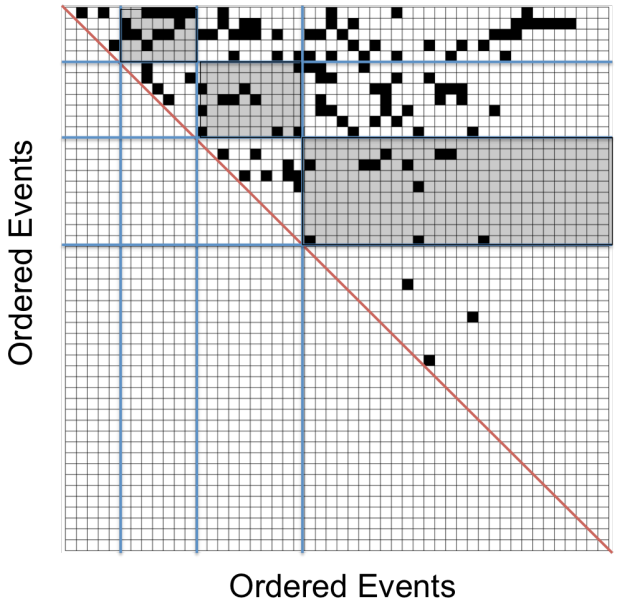
Event Grouping

Given the TOE graph, a natural follow-up question is whether there are groups of events that succeed certain events, precede other events, but show no specific relations amongst themselves, i.e. – are there groups of events that are indistinguishable by their timing?

To address this question, and given the total order described in the previous section, we wish to partition the events to groups and optimize the total weight of edges between subsequent groups. Visually, if the edge matrix is the following binary matrix (this is easily extended to non-binary matrices, or weighted edges):



We are looking for a partition of the events (collection of blue lines) that maximizes the sum of the highlighted rectangles:



This problem can be solved using a simple dynamic programing. By iteratively answering the question: "which is the optimal partition up to index i", and selecting the i for which the maximal-valued partition is obtained. Note however, that the conjoining of two sub-solutions also requires us to know the size of the last set in the optimal sub partition. Since every partition must have a final set, and the order is fixed, this strategy searches the partition-space exhaustively for the optimal solution in $O(n^3)$ time, which is reasonable for our problem size ($n \leq 200$).

Supplemental References

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368-373.

Dai, J., Hyland, E.M., Yuan, D.S., Huang, H., Bader, J.S., and Boeke, J.D. (2008). Probing nucleosome function: a highly versatile library of synthetic histone H3 and H4 mutants. Cell *134*, 1066-1078.

Dion, M.F., Kaplan, T., Kim, M., Buratowski, S., Friedman, N., and Rando, O.J. (2007). Dynamics of replication-independent histone turnover in budding yeast. Science *315*, 1405-1408.

Egelhofer, T.A., Minoda, A., Klugman, S., Lee, K., Kolasinska-Zwierz, P., Alekseyenko, A.A., Cheung, M.S., Day, D.S., Gadel, S., Gorchakov, A.A.*, et al.* (2011). An assessment of histone-modification antibody quality. Nat Struct Mol Biol *18*, 91-93.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell *11*, 4241-4257.

Kim, T.S., Liu, C.L., Yassour, M., Holik, J., Friedman, N., Buratowski, S., and Rando, O.J. (2010). RNA polymerase mapping during stress responses reveals widespread nonproductive transcription in yeast. Genome Biol *11*, R75.

Ni, T., Corcoran, D.L., Rach, E.A., Song, S., Spana, E.P., Gao, Y., Ohler, U., and Zhu, J. (2010). A paired-end sequencing strategy to map the complex landscape of transcription initiation. Nat Methods *7*, 521-527.

Raghuraman, M.K., Winzeler, E.A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D.J., Davis, R.W., Brewer, B.J., and Fangman, W.L. (2001). Replication dynamics of the yeast genome. Science *294*, 115-121.

Weiner, A., Chen, H.V., Liu, C.L., Rahat, A., Klien, A., Soares, L., Gudipati, M., Pfeffner, J., Regev, A., Buratowski, S.*, et al.* (2012). Systematic dissection of roles for chromatin regulators in a yeast stress response. PLoS Biol *10*, e1001369.