PNAS
www.pnas.org

Supplementary Information for

Chromatin conformation remains stable upon extensive transcriptional changes driven by heat shock

Judhajeet Ray[1*], Paul R. Munn[2*], Anniina Vihervaara[1], James J Lewis[2], Abdullah Ozer[1†], Charles G. Danko[2†], and John T. Lis[1†]
[1]Department of Molecular Biology and Genetics, [2]Baker Institute for Animal Health, Cornell University, Ithaca, NY 14853

*These authors contributed equally to this work

[†]To whom correspondence should be addressed. Email: ao223@cornell.edu OR cgd24@cornell.edu OR jtl10@cornell.edu

This PDF file includes:

Supplementary text
Figures S1 to S8
Tables S1 to S2
Legends for Datasets S1 to S2
SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S2

www.pnas.org/cgi/doi/10.1073/pnas.1901244116

Supplementary Information text

MATERIALS AND METHODS

Heat shock treatments & formaldehyde crosslinking

Heat shock (HS) for human K562 chronic myelogenous leukemia cells were done according to the protocol described previously (1). The cells were grown in RPMI medium with 10% FBS at 37°C. Prior to treatment cells were concentrated in 10 ml medium. Cells were incubated at 37°C for non-heat shock (NHS) or at 42°C for HS in water bath for 30 min. After treatment cells were centrifuged for 5 min at 4°C and the supernatant was discarded. Cells were resuspended in 1x PBS and crosslinked with formaldehyde to a final concentration of 1% for 10 min at room temperature with occasional mixing. Crosslinking was quenched by the addition of 200 mM glycine for 5 min at room temperature with mixing. Cells were centrifuged at 500 x g for 5 min at 4°C and the medium was discarded. The cells were washed once with 1x PBS, pellets were flash frozen in liquid nitrogen and stored at -80°C.

Drosophila S2 cells were grown in M3+BPYE medium with 10% FBS at 25°C. Cells were transferred to a shaking water bath maintained at room temperature for NHS or at 36.5°C for HS. Simultaneously, an equal volume of medium (without FBS), kept at room temperature or at 48°C was added into NHS or HS cells respectively. Cells were then incubated for 20 min. Cells were centrifuged at 500 x g at 4°C for 5 min. The medium was removed and cells were resuspended in 1x PBS. Crosslinking was done with the addition of formaldehyde to a final concentration of 1%, and incubated for 10 min at room temperature with occasional mixing. Glycine was added to a final concentration of 147 mM, and incubated at room temperature for 5 min with mixing. Cells were centrifuged and the supernatant was discarded. The cells were washed once with 1x PBS, pellets were flash frozen in liquid nitrogen and stored at -80°C.

Hi-C library preparation

We used 58 million S2 or 5 million K562 cells for in situ Hi-C (2). The frozen crosslinked cell pellets were thawed on ice and resuspended in 250 µL ice-cold Hi-C lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% NP40) with Protease inhibitor cocktail (Thermo). Cells were incubated on ice for 30 min, centrifuged and washed once with 500 µL ice-cold Hi-C lysis buffer. The pellet was resuspended in 50 µL of 0.5% SDS and incubated at 62°C for 7 min followed by addition of 145 µL water and 25 µL of 10% Triton-X-100 and incubated at 37°C for 15 min. Finally, 25 µL of 10x NEBuffer2 and 5 µL (125 units) of MboI (NEB) were added to the mixture. The sample was digested at 37°C overnight with rotation. The sample was incubated at 62°C for 20 min to inactivate MboI and then cooled to room

temperature. Biotin fill-in of digested ends was done by adding 1.5 µL each of 10 mM dCTP, dGTP, dTTP, 37.5 µL of 0.4 mM Biotin-14-dATP (Thermo) and 8 µL (40 units) of Klenow polymerase (NEB). The reaction was incubated at 37°C with rotation for 90 min. Ligation was performed by addition of 100 µL 10% Triton-X-100, 120 µL of 10x T4 DNA ligase buffer (NEB), 12 µL of 10 mg/ml BSA (NEB), 5 µL (2000 units) of T4 DNA ligase (NEB) and 663 µL of water. The ligation mixture was incubated at room temperature for 4 h with rotation. To reverse the crosslinks, 50 µL of 20 mg/ml Proteinase K (NEB) and 120 µL of 10% SDS were added to the sample followed by an incubation at 55°C for 30 min. To this mixture 130 µL of 5 M NaCl was added and incubated at 68°C overnight. The reactions were cooled to room temperature and the sample was purified by 1.6x volume of 100% ethanol and 0.1x volume of 3 M Sodium acetate, pH 5.2 followed by incubation at -80°C for 15 min. The sample was pelleted by spinning at 20,000 x g at 4°C and washed twice with 70% ethanol. The pellet was resuspended in 110 µL of 10 mM Tris-HCl, pH 8.0 and incubated at 37°C for 15 min to dissolve. The purified sample was sonicated using a Bioruptor Diagenode sonicator at low setting, with 30 second on and 90 second off for 20 min in an ice-cold water bath at 4°C. The sonicated sample was then treated with 2 µL RNase A/T1 cocktail (Thermo) for 30 min at 37°C. The DNA was cleaned up using Qiaquick PCR purification kit (Qiagen) and the eluate volume was brought up to 300 µL with 10 mM Tris-HCl, pH 8.0. Biotin pull down was done with 150 µL of Dynabeads MyOne Streptavidin T1 (Thermo) that was washed with 400 µL of 1x Tween wash buffer (5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween-20). Washed beads were resuspended in 300 µL of 2x Binding and wash buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 2 M NaCl) and added into 300 µL of DNA sample. Binding was done at room temperature for 15 min with rotation. Beads were washed twice with 1x Tween wash buffer, transferred into a new tube and heated at 55°C for 2 min with shaking in a thermomixer. Beads were then washed once with 100 µL of 1x NEB T4 DNA ligase buffer and transferred into a new tube. End repair was done by adding 88 µL of 1x NEB T4 DNA ligase buffer, 2 µL of 25 mM dNTP, 5 µL (50 units) of T4 PNK (NEB), 4 µL (12 units) of T4 DNA polymerase (NEB) and 1 µL (5 units) of Klenow polymerase (NEB). The mixture was incubated at room temperature for 30 min. Beads were washed with 1x Tween wash buffer as described previously and once with 100 µL of 1x NEBuffer2, and transferred into a new tube. A-tailing was done by adding 90 µL of 1x NEBuffer2, 5 µL of 10 mM dATP and 5 µL (25 units) of Klenow (3'-5' exo-) polymerase (NEB). The reaction was incubated at 37°C for 30 min. Beads were washed with 1x Tween wash buffer just as before and once with 100 µL of 1x NEB T4 DNA ligase buffer, and transferred to a new tube. Adaptor ligation was done by adding 50 µL of 1.1x NEB T4 DNA ligase buffer, 2 µL of 3 µM Truseq/Universal indexed adaptor and 3 µL (1200 units) of T4 DNA ligase (NEB).

Ligation was done at room temperature for 2 h. Beads were washed with 1x Tween wash buffer as described previously and test PCR was performed with P5 and P7 primers to determine the optimal cycle number for library amplification. Final PCR was done with 90% of the sample for 8-10 cycles. The final amplified library was purified with Qiaquick PCR purification kit (Qiagen) and sequenced using Illumina Nextseq 500.

Contact map similarity

Visualizations of the contact maps, PRO-seq data, and ChIP-seq data were produced using Juicebox (2), HiGlass (3), and pyGenomeTracks (4).

To measure contact map similarity beyond our initial visual comparison, we used HiCRep to obtain a stratum-adjusted correlation coefficient (SCC) (5). This was done for all chromosomes in human and *Drosophila*, in both cases using a bin size of 10 Kb and maximum interaction distance set to 5 Mb. For human, a smoothing parameter of 12 was used; for *Drosophila,* 8 was used, as recommended by the HiCRep software.

To determine compartment type (active or inactive) and compartmentalization strength, we computed the first principal component (PC) of the Pearson's correlation matrix of the observed contact map / expected contact map, across the entire genome. Pearson correlation matrices were computed using the Pearsons tool, and the PC computed using the eigenvector tool (2), both using a bin size of 50 Kb on KR normalized datasets. For *Drosophila* data, we used the “-p” option to the eigenvector tool to ignore sparsity when calculating the PC.

Typically, open / active compartments are defined as having positive PC1 values. However, we noted that the sign of the PC was reversed on some chromosomes. In order to confirm each compartment call, we used Pol II occupancy, as indicated by our PRO-seq data and the ChIP-seq signal for H3K36me3 (ENCODE) in the same region as markers for open chromatin to switch the PC sign when necessary. In all cases, we defined positive PC1 scores (the A compartment) as the compartment that correlated with active transcription and H3K36me3.

Classification of up- and down-regulated genes was conducted with DESeq2 (6) and is described in details in (7). Only high confidence (HC) calls, p-value < 0.05 and fold change > 1.25, were used in downstream analyses (i.e. Regulation upon 30 minutes of HS UpHC or DownHC). The high-confidence, HS up-regulated genes were subsequently split into HSF1-dependent and HSF1-independent using HSF1-dependent gene calls generated in (7).

TAD-separation score was determined across the entire genome using the HiCExplorer hicFindTads tool (4). Default values were used for the 'correctForMultipleTesting' and 'thresholdComparisons' parameters ('fdr' and '0.01'

respectively), and 'minDepth', 'maxDepth', and 'step' were set to 30 Kb, 100 Kb, and 10 Kb respectively.

Contact prediction modeling

We used gradient boosted trees implemented in scikit-learn's ADABoostClassifier (8). Genes were randomly split into training and test datasets (80% of the data used for the training set and 20% held out for testing) and the classifier was trained on various combinations of features (or models) from this data (see below). All models were trained to discriminate between HSF1-dependent and HSF1-independent up-regulated genes on the basis of the nearest HSF1 binding site. A precision / recall curve was produced based on the correct and incorrect predictions made by the classifier and the area under the curve (AUC) calculated to assess of how well the classifier performed. This process of randomly splitting the data followed by training a classifier was performed 1000 times to get a spread of AUC values for each model. Values were plotted on a boxplot (Fig. 5B). A Wilcoxon rank-sum test was used to compare how each model performed relative to the others. Multiple models were tested in order to determine the predictive value of various features within our dataset. Results for the following models are shown in figure 5B:

Control: Bypasses the classifier predictions and represents the results obtained by random chance.

Dist / PS for closest peak: Uses the distance from the gene TSS to the closest peak and the strength of that peak.

Dist / PS / CF: Similar to the previous model with the addition of the contact frequency to the closest peak.

We defined HSF1 dosage as:

$$Dose \equiv \sum_{i \in n} S_i \frac{1}{1 + e^{-a(C_i - b)}} t_i$$

where *i* represents the HSF1 binding site from {1..n} that are within 1 Mb of each gene's TSS; S_i represents the fold enrichment of HSF1 signal in binding site *i* over input control called by MACS 1.4 (9); t_i was set as a 1 if peak *i* intersected a dREG transcription initiation region (10) and otherwise was set to a value of *z* (see below). C_i was the contact frequency between HSF1 binding site *i* and the TSS; *a*, *b*, and *z* were free parameters of the model that were optimized to maximize auPRC on a training set of genes. In all cases, the training and test sets of genes were the same as used for the gradient boosted trees (described above). We noted that optimizing *a*, *b*, and *z* simultaneously using L-BFGS or conjugate gradients resulted in poor performance, in which the fitted values were nearly unchanged from the starting values. Therefore, we used Brent's method

(11) to identify the values of *a*, *b*, or *z* that maximized auPRC while holding the other values constant. We performed three rounds of optimization for each parameter. We bounded values of $a = [0, 1]$, $b = [0, 500]$, and $z = [0,1]$.

Significant interactions between HSF1 binding sites and gene TSSs were called as follows: A local background model was calculated from the empirical distribution of the distance between aligned read pairs where at least one read aligns within 250 Kb of the HSF1 ChIP-seq peak tested. A 5 Kb window centered on the HSF1 peak was used to capture all reads falling into the “bait” region, and for each TSS within 250 Kb of the HSF1 peak a second 5 Kb window centered on the TSS (the “prey” region) was used to capture the number of Hi-C contacts between the HSF1 binding site and the TSS. To calculate the expected number of contacts, we took the integral of the empirical distribution over the minimum and maximum distances between the bait and prey windows. This frequency was then multiplied by the total number of read pairs captured by the bait window to determine the expected number of contacts. Fisher’s exact test was used to compare observed to expected contacts and determine significance. Subsequent analyses were performed using a p-value threshold of $p = 0.01$. Conservatively assuming all significant interactions between HSF1 binding sites and HSF1-unregulated genes are false positives, the same $p = 0.01$ gave a false positive interaction call ceiling of 6%.

When calling for significant interactions between HS up-regulated dTREs and gene TSSs the abovementioned method was followed by changing the “bait” to HS up-regulated dTREs.

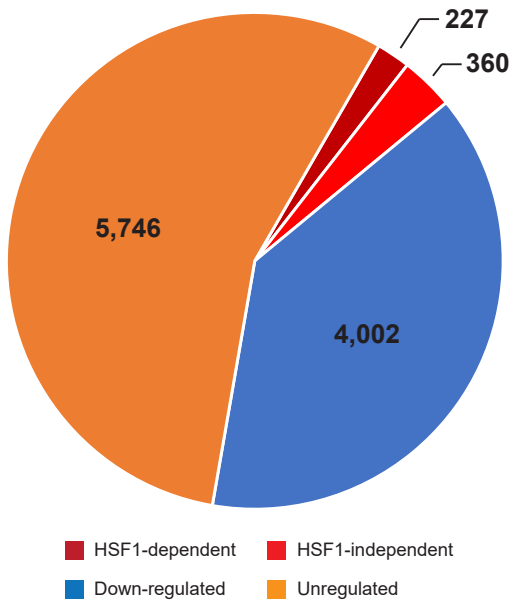

Fig. S1. Distribution of genes between various classes (HSF1-dependent and HSF1-independent up-regulated, HS down-regulated, and HS unregulated).

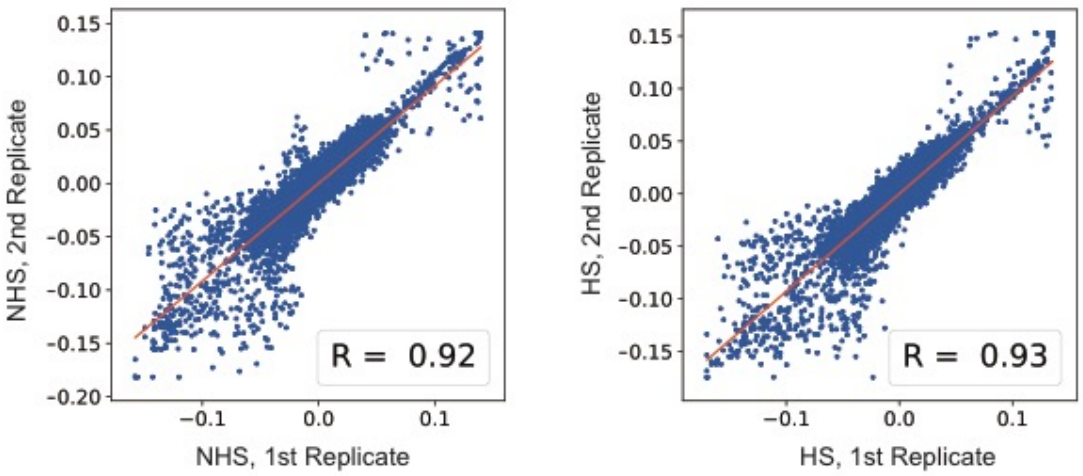

Fig. S2. Correlation between the strength of compartment calls for replicates from the same condition, genome-wide in K562.

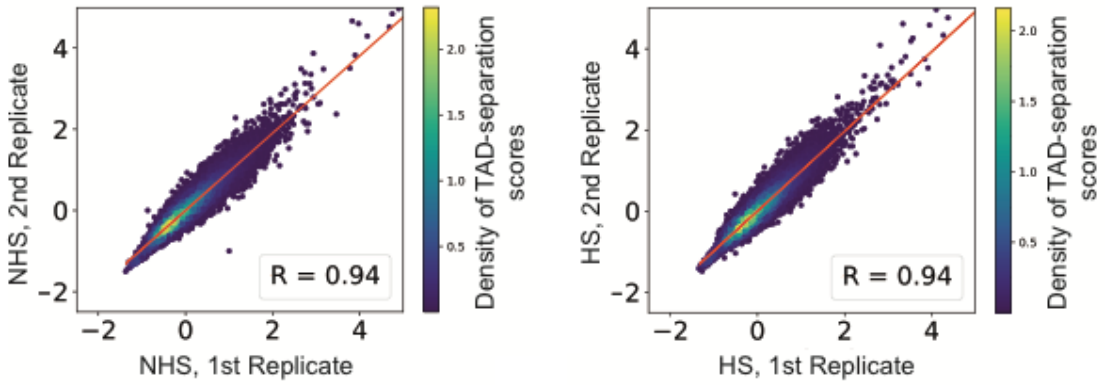

Fig. S3. Correlation of TAD-separation scores for replicates from the same condition, genome-wide in K562.

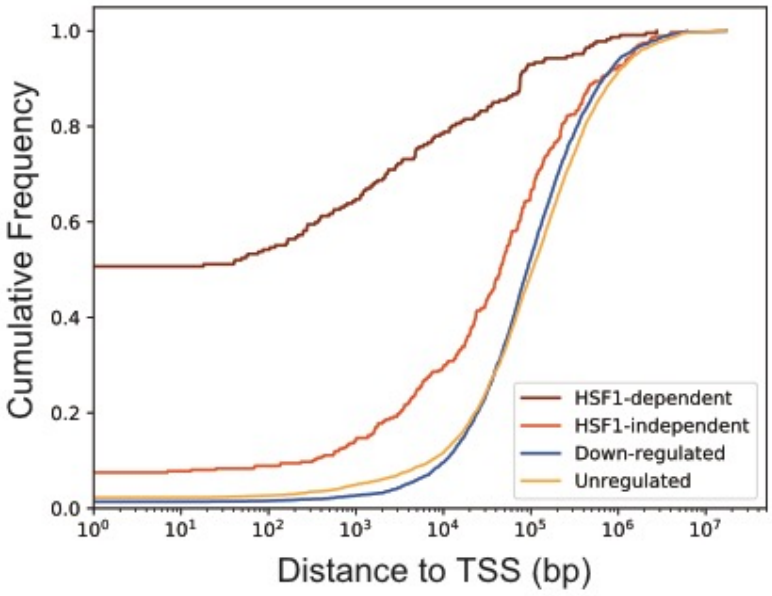

Fig. S4. Cumulative frequency of the distances in base pairs between the peaks of HSF1 binding sites and TSS of HSF1-dependent, HSF1-independent, down-regulated and unchanged genes.

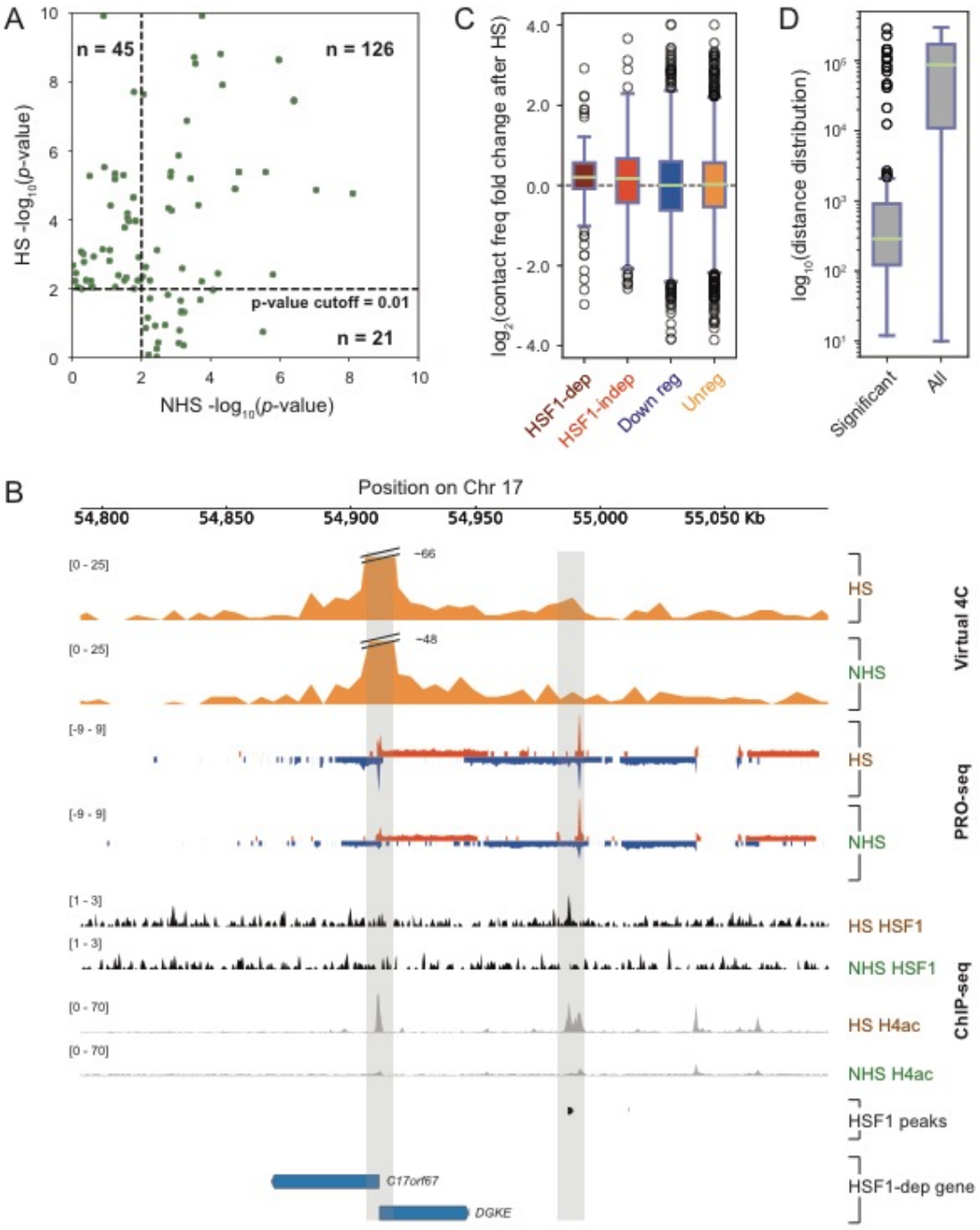

Fig. S5. A) Looping interactions between HSF1 binding sites and HSF1-dependent gene promoters for HS and NHS conditions (shown in Fig. 4A) rescaled to highlight the interactions below the cutoff (p = 0.01).
B) Virtual 4C plot showing contacts between *DGKE* TSS and its nearest HSF1 binding site (grey bars). PRO-seq tracks are included to show changes in transcription for this gene upon HS. ChIP-seq tracks of HSF1 and H4ac show HSF1 binding and accompanying active chromatin status respectively upon HS.

C) Change in observed/expected looping interactions for HS and NHS Hi-C samples, for HSF1-dependent up-regulated, HSF1-independent up-regulated, down-regulated, and unregulated genes with HS up-regulated dTREs, using a cutoff at which <10% of the unregulated interaction calls were significant.
D) Distance distribution of the significant and all looping interactions between HSF1 binding sites and HSF1-dependent gene promoters.

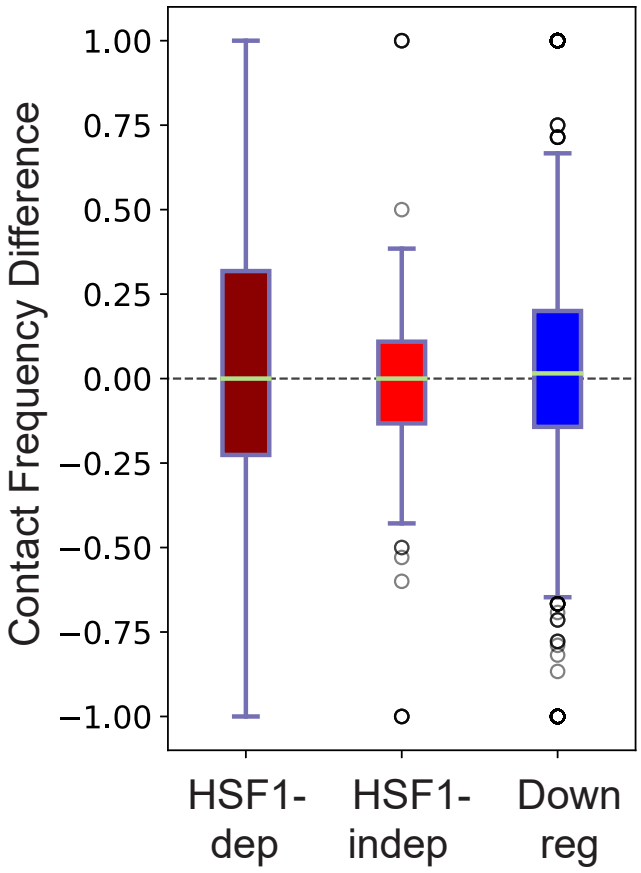

Fig. S6. Difference in contact frequencies between gene TSSs and distal (≥ 10 Kb) enhancer sites for HSF1-dependent up-regulated, HSF1-independent up-regulated, and HS down-regulated genes under NHS and HS conditions. Contact frequencies were calculated for HSF1-dependent genes between TSS and HSF1 binding sites; for HSF1-independent genes between TSS and active, up-regulated transcriptional regulatory elements; for down-regulated genes between TSS and active, down-regulated transcriptional regulatory elements.

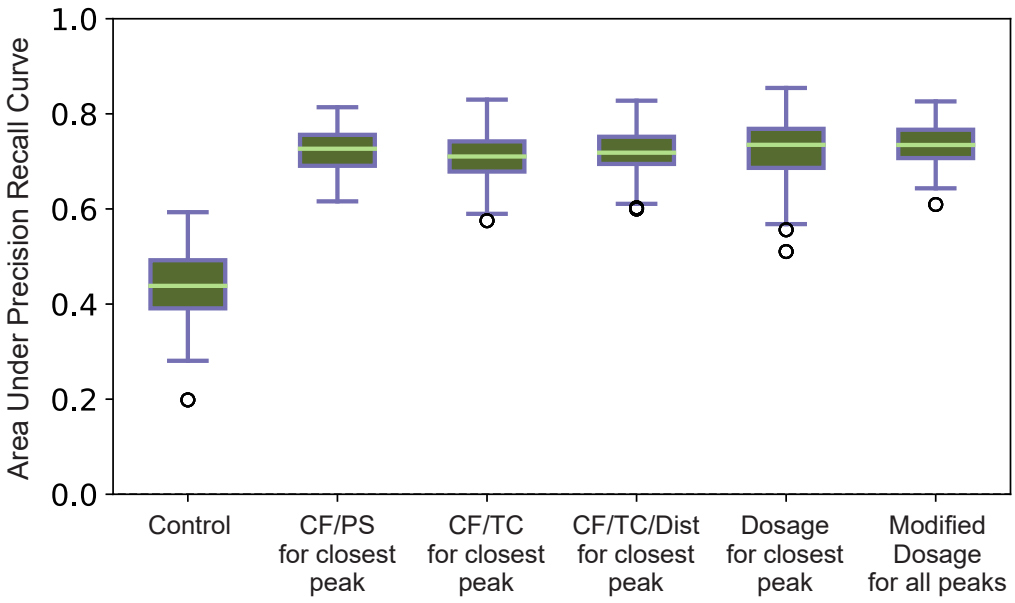

Fig. S7. Boxplots showing spread of area under the precision recall curve (auPRC) results for 1000 iterations of classifiers not included in Fig. 5B. From left to right, **Control**: results obtained by randomly selecting the gene class; **CF/PS for closest peak**: contact frequency and peak strength (fold enrichment of HSF1 binding signal) of the closest HSF1 binding site; **CF/TC for closest peak**: contact frequency and transcribed count (PRO-seq reads) at the closest HSF1 binding site for NHS data; **CF/TC/Dist for closest peak**: contact frequency, linear distance, and transcribed count (PRO-seq reads) at the closest HSF1 binding site for NHS data; **Dosage for closest peak**: scaled contact frequency multiplied by peak strength for closest HSF1 binding site; **Modified dosage for all peaks**: scaled contact frequency multiplied by peak strength for all transcribed HSF1 binding sites within 1 Mb window flanking each gene TSS.

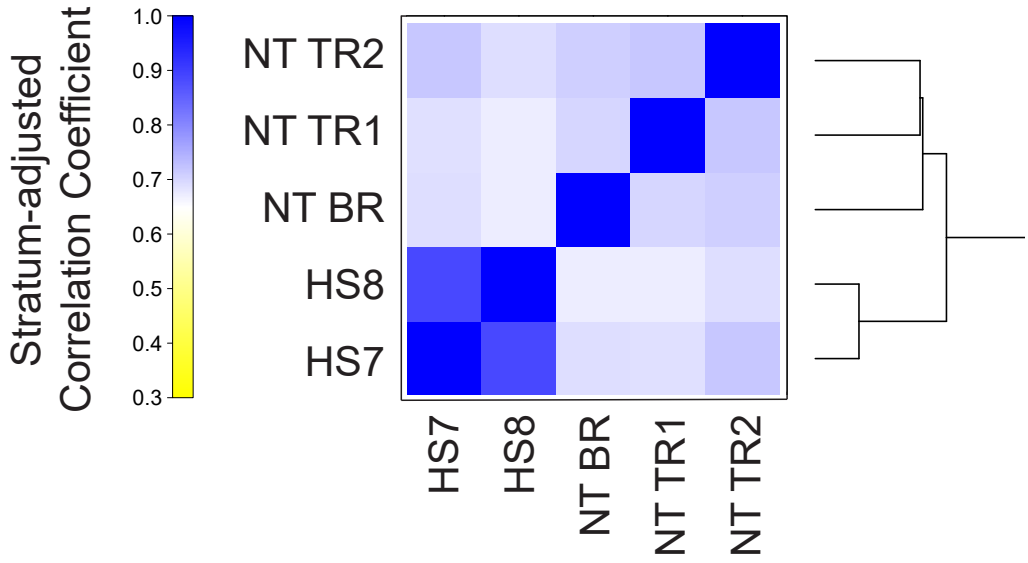

Fig. S8. Stratum-adjusted correlation coefficients for heat shocked and non-treated (NT/NHS) replicates. Data for HS is from (12) and non-treated conditions is from (13).

Table S1. Hi-C sequencing statistics

K562 Datasets	Sequenced Read Pairs	Hi-C Contacts
HS Rep 1	86,303,516	37,471,475
HS Rep 1 resequenced	51,239,864	24,312,467
HS Rep 2	127,112,460	50,286,960
HS Rep 2 resequenced	78,659,141	34,899,255
NHS Rep 1	86,454,136	37,388,359
NHS Rep 1 resequenced	135,196,055	63,130,323
NHS Rep 2	128,491,399	50,466,641
NHS Rep 2 resequenced	48,902,638	21,528,585

S2 Datasets	Sequenced Read Pairs	Hi-C Contacts
HS Rep 1	156,103,131	62,257,766
HS Rep 2	151,421,737	59,528,855
NHS Rep 1	148,391,404	61,410,790
NHS Rep 2	124,408,163	49,700,701

.

Table S2. Correlation between Hi-C datasets

K562 Datasets	Stratum-adjusted Correlation Coefficient
HS 1st rep vs. HS 2nd rep	0.905
HS 1st rep vs. NHS 1st rep	0.905
HS 1st rep vs. NHS 2nd rep	0.896
HS 2nd rep vs. NHS 1st rep	0.899
HS 2nd rep vs. NHS 2nd rep	0.915
NHS 1st rep vs. NHS 2nd rep	0.898

S2 Datasets	Stratum-adjusted Correlation Coefficient
HS 1st rep vs. HS 2nd rep	0.943
HS 1st rep vs. NHS 1st rep	0.955
HS 1st rep vs. NHS 2nd rep	0.928
HS 2nd rep vs. NHS 1st rep	0.971
HS 2nd rep vs. NHS 2nd rep	0.979
NHS 1st rep vs. NHS 2nd rep	0.972

Dataset S1. List of genes and their transcriptional status in human K562 cells upon HS

See the text file SD1_geneRegulation_upon_30minHS_K562_scr

chr: chromosome, txStart: transcript start, txEnd: transcript end, geneName: the name of the gene, refGeneName: transcript identification, UpHC: heat-induced with high confidence (counted as up-regulated in this study), UpLC: heat-induced with low confidence, DownHC: heat-repressed with high confidence (counted as down-regulated in this study), DownLC: heat-repressed with low confidence, UnReg: unchanged transcription, UnExp: very low gene body transcription.

Dataset S2. List of non-transcribed genes in NHS that locate to B compartment

See the text file SD2_unExp_geneList_K562_NHS_B_comp_only

chr: chromosome, txStart: transcript start, txEnd: transcript end, geneName: the name of the gene, refGeneName: transcript identification.

References

1. A. Vihervaara *et al.*, Transcriptional response to stress is pre-wired by promoter and enhancer architecture. *Nat Commun* **8**, 255 (2017).
2. S. S. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).
3. P. Kerpedjiev *et al.*, HiGlass: Web-based Visual Comparison And Exploration Of Genome Interaction Maps. *bioRxiv* (2017).
4. F. Ramirez *et al.*, High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun* **9**, 189 (2018).
5. T. Yang *et al.*, HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res* **27**, 1939-1949 (2017).
6. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
7. A. Vihervaara *et al.*, Stress-Induced Transcriptional Memory Accelerates Promoter-Proximal Pause-Release and Decelerates Termination over Mitotic Divisions. *bioRxiv* 10.1101/576959, 576959 (2019).
8. F. Pedregosa *et al.*, Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res* **12**, 2825-2830 (2011).
9. J. Feng, T. Liu, B. Qin, Y. Zhang, X. S. Liu, Identifying ChIP-seq enrichment using MACS. *Nat Protoc* **7**, 1728-1740 (2012).
10. C. G. Danko *et al.*, Identification of active transcriptional regulatory elements from GRO-seq data. *Nat Methods* **12**, 433-438 (2015).
11. R. P. Brent, An algorithm with guaranteed convergence for finding a zero of a function. *The Computer Journal* **14**, 422-425 (1971).
12. L. Li *et al.*, Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. *Mol Cell* **58**, 216-231 (2015).
13. C. Hou, L. Li, Z. S. Qin, V. G. Corces, Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol Cell* **48**, 471-484 (2012).