Cell cycle arrest explained the observed bulk 3D genomic alterations in response to long-term heat shock in K562 cells

Bingxiang Xu[1,2,3,#,*], Xiaomeng Gao[1,2,#], Xiaoli Li[1,2,4], Yan Jia[1], Feifei Li[1,5,*], Zhihua Zhang[1,2,*]

1, Beijing Institute of Genomics, Chinese Academy of Sciences, and China National Center for Bioinformation No.1 Beichen West Road, Chaoyang District, Beijing 100101, China.

2, School of Life Science, University of Chinese Academy of Sciences, Beijing, China.

3, School of Kinesiology, Shanghai University of Sport, Shanghai, China.

4, Department of Cell Biology and Genetics, Core Facility of Developmental Biology, Chongqing Medical University, Chongqing 400016, China.

5, Division of Cell, Developmental and Integrative Biology, School of Medicine, South China University of Technology, Guangzhou 510006, China.

#: These authors contributed equally.

*: Correspondence should be sent to xubingxiang@sus.edu.cn (XBX); liff@scut.edu.cn (LFF) or zhangzhihua@big.ac.cn (ZZH).

Supplemental Material

Stability of the loop domains during heat shock process

To further indicate the stability of TADs boundaries, we specially analyzed the loop domains, which were domains with boundaries connected by a loop (Rao et al. 2014; Sanborn et al. 2015). The position of loop domains also prone to stable during heat shock process. For example, we compared the loop domain boundaries between NHS and LHS in the mixed cells, and found even after long term heat shock, boundaries in one condition remains tended to be boundaries in the other (Supplemental Fig. S2C). Moreover, loop domain regions were significantly overlapped between the two conditions (25.20% of the genome regions identified as loop domain in one condition were identified as loop domains by both conditions, $p < 0.001$, permutation test). This stability of loop domains was also observed in the G2/M cells (Supplemental Fig. S7C).

Loops were stable when using a reference loop list

To further show the stability of the loops, we merged all reads from the three conditions in the mixed cells and called 17069 loops as reference loops. We found that most of these reference loops consistently have higher-than-expected contact frequency (observed / expected) in the three conditions and the cosine similarity scores for the contact frequencies of these reference loops were all larger than 0.96 in all comparisons (Supplemental Fig. S2F and G), which further indicating the loop changes are minor.

Mixture of different phases of cells may result in larger compartment score than its components

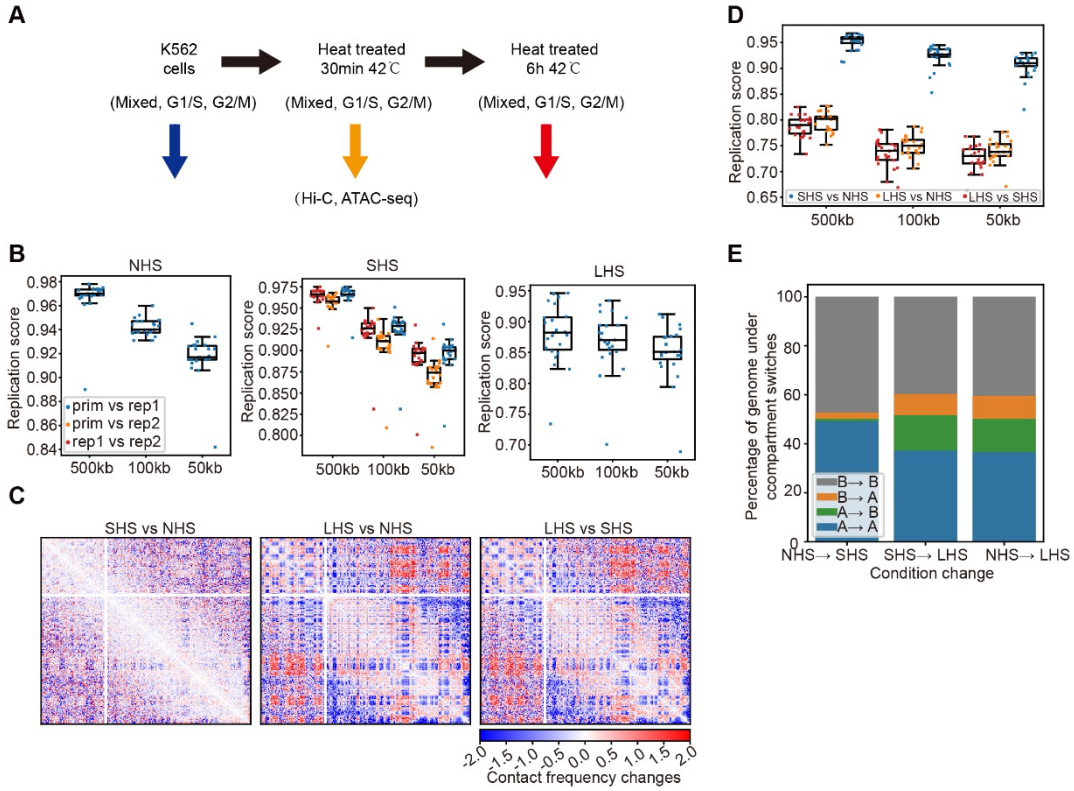
How the compartment score may react to changes of population architecture, i.e., whether linear, monotonous or convex, remains elusive. Therefore, we decided to examine it numerically with real data. We generated a series of mixed matrices with G1/S and G2/M cells by randomly sampling reads from the two libraries. The fraction of G1/S cells was arbitrarily set as 0%, 10%, 20%, …, 90% and 100%. Interestingly, in 11 out of the 23 chromosomes, we found that the maximum compartment score of mixed matrices is larger than that in either G1/S or G2/M. For example, the compartment scores of chr18 in G1/S and G2/M cells are 0.25 and 0.26, respectively, while the maximum compartment score of mixed population is 0.28 (> 0.26) when the proportion of G1/S cells is 40% (Supplemental Fig. S10). Thus, it is not surprising that we see the larger-than-subcomponent compartment score in mixed cells.

Accuracy of the TADs called by DeDoc in this study

To verify the insulation level of the TAD boundaries called in this study, we directly compared the boundaries to its nearest local minima of the insulation score profile. We found that more boundaries were predicted close to these minima loci with our deDoc than public tools did. For example, 52.02% and 44.81% of the TADs boundaries were predicted

no larger than 50kb, by our deDoc in our data and Arrowhead in their origin paper (Rao et al. 2014), respectively. The average distances of our TAD boundaries were also significantly smaller than those called by Arrowhead (Supplemental Fig. S13).

Supplemental Figures



Supplemental Fig. S1. Chromatin conformation of K562 cells in the process of HS.

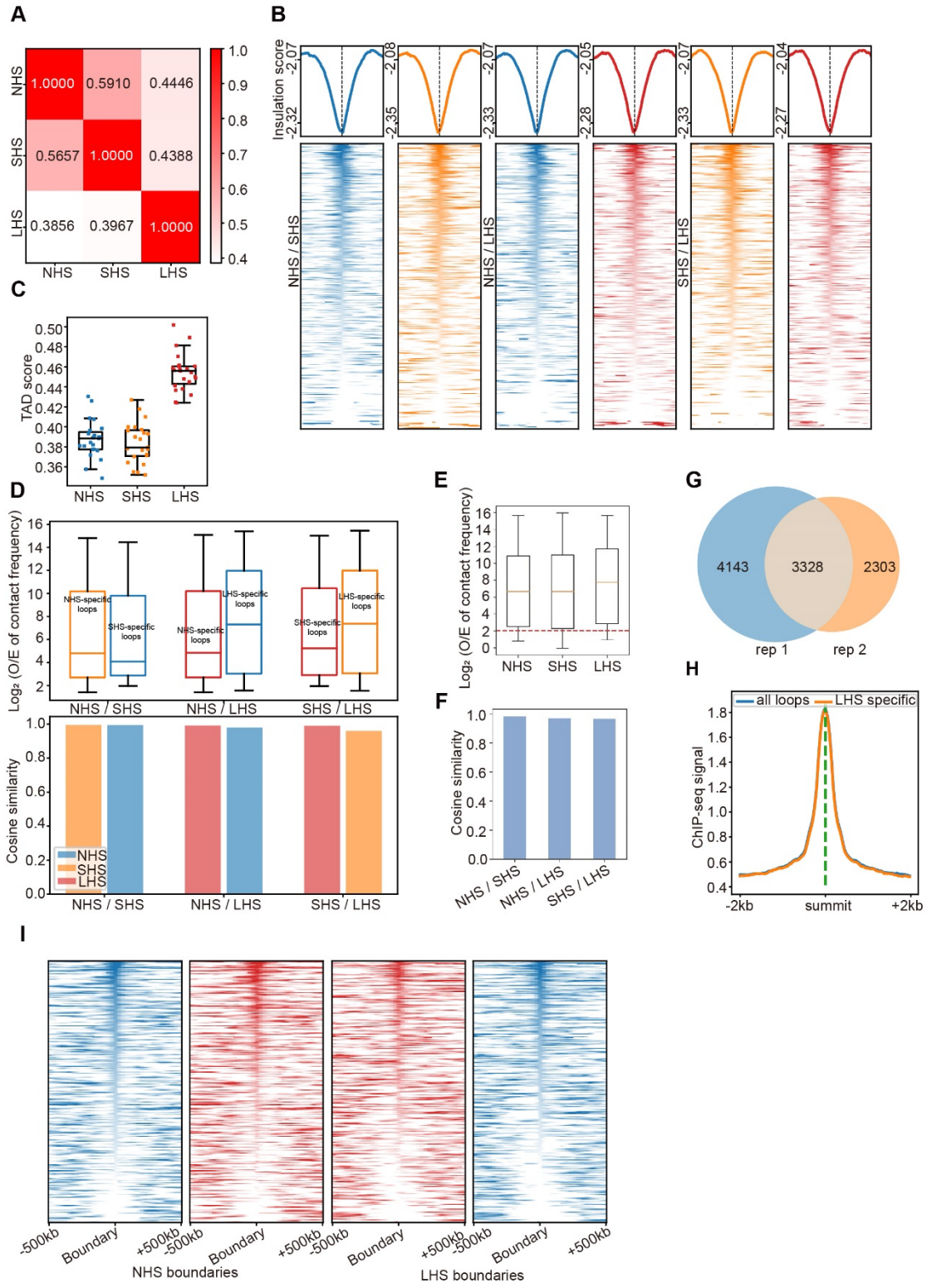
(A) Sketch map showing experimental design.

(B) GenomeDisco scores showing the reproducibility of the Hi-C libraries.

(C) Fold changes of contact frequencies (observed / expected KR normalized) for Chr6 between any pair of the three conditions.

(D) GenomeDisco scores among different conditions.

(E) Proportions of genome regions which switched compartments after SHS and LHS.



Supplemental Fig. S2. Changes of TADs and loops during HS in the mixed cells.

(A) Changes in TAD boundary position during the process of HS. The values in row i and column j indicate the proportion of TAD boundaries in i which distanced no more than 1 bin with a boundary in condition j.

(B) The average and heatmap of insulation scores (±500kb) around condition-specific TAD boundaries in conditions not called. The two conditions compared were marked on the left, and

colors indicated the condition in which the insulation scores were calculated (NHS: blue, SHS: brown, LHS: red). For example, the left blue heatmap showed the insulation score calculated in NHS for SHS-specific TAD boundaries.

(C) Insulation score profiles around boundaries of loop domains in NHS and LHS for mixed cells. The colors coded the conditions from which the profiles were calculated (blue: NHS, red: LHS).

(D) TAD strength scores (inter-TAD contacts/intra-TAD contacts) of the three conditions.

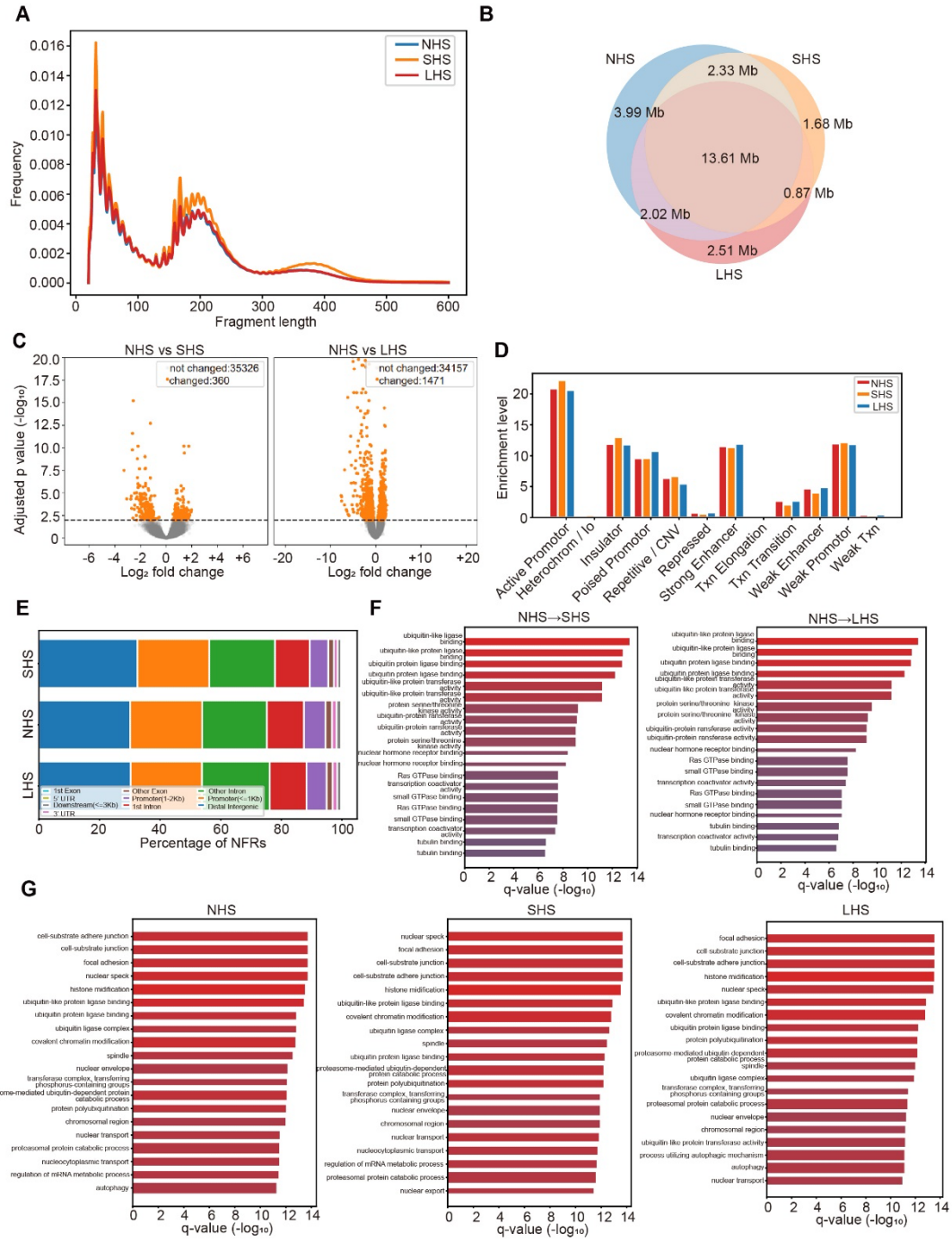
(E) Comparison of contact frequencies of condition-specific loops in mixed cells. When comparing the two conditions labeled in the X-axis, the upper boxplot showed the contact frequencies calculated in color-coded condition for specific loops called in the other condition. For example, the left orange boxplot showed the contact frequencies in SHS condition for NHS-specific loops. The lower panel showed the cosine similarity scores between the two labeled conditions in X-axis for loops specific to the color-coded condition. For example, the left orange bar showed the similarity between NHS and SHS for SHS-specific loops.

(F) Contact frequencies (observed / expected) of the reference loops in the three conditions.

(G) Cosine similarity scores of contact frequencies in reference loops between any two conditions.

(H) The proportion of shared loops between biological replicates in the public GM12878 datasets (Rao et al. 2014).

(I) CTCF ChIP-seq signals in the anchors of LHS-specific loops and all combined loops in NHS condition. Binding affinities were calculated from the ChIP-seq data in ENCODE (The ENCODE Project Consortium 2012).



Supplemental Fig. S3. Stability of chromatin states before and after HS, as measured by ATAC-seq.

(A) Fragment length distributions of ATAC-seq at the three conditions.

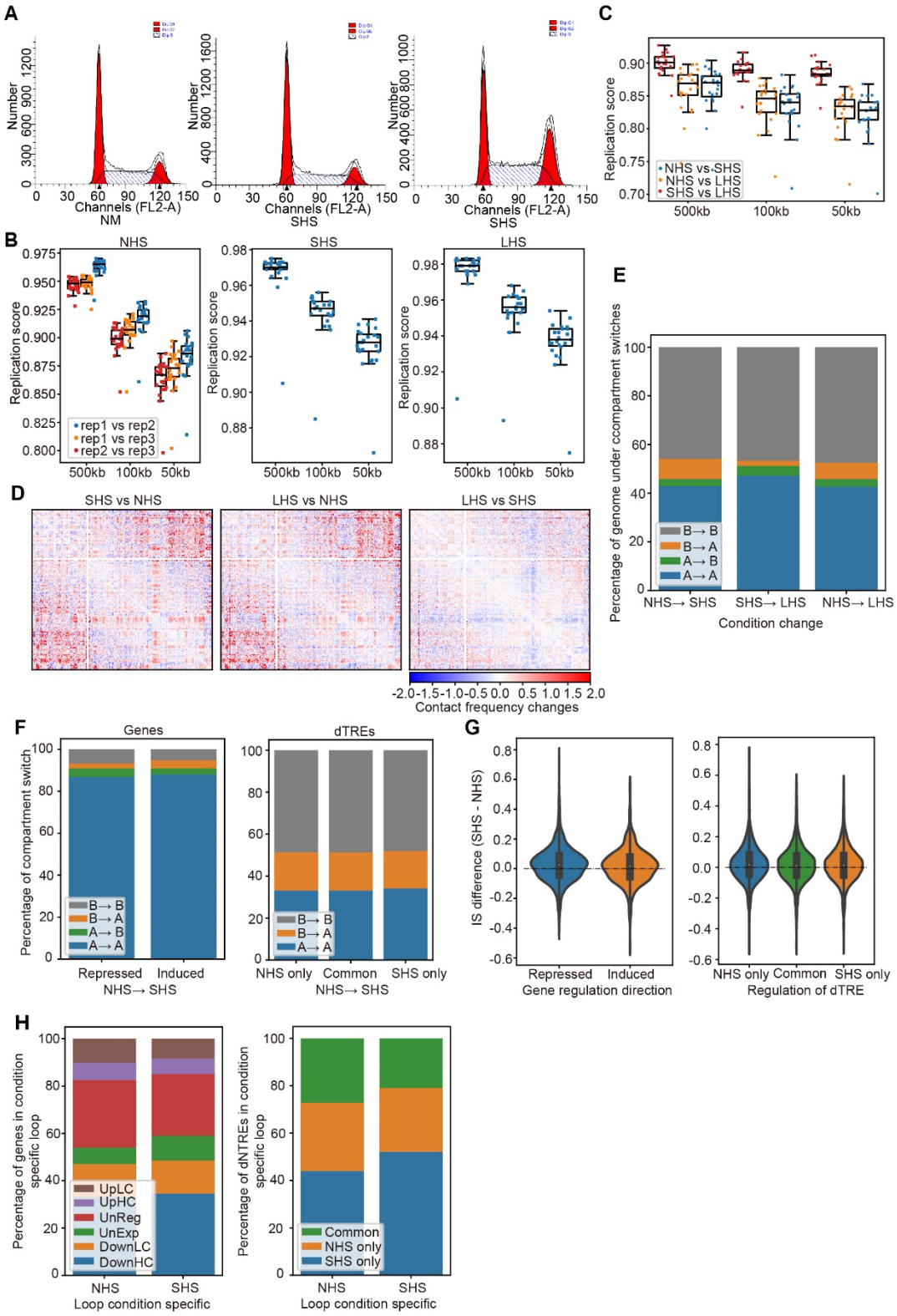
(B) Comparison of genome lengths identified as MACS2 peaks in the ATAC-seq of the three conditions.

(C) Volcano plot showing the results of differential accessibility analysis in MACS2 peaks before and after HS.

(D and E) Distributions of NFRs in different chromatin states and genetic regions.

(F) GO terms enriched in genes specifically accessible (with TSSs covered by the NFRs) after short-term (left panel) and long-term (right panel) HS.

(G) GO terms enriched by genes containing NFR in the TSS region at each condition (NHS, SHS and LHS from left to right).



Supplemental Fig. S4. Chromatin conformation in G1/S cells remains largely intact after long-term HS.

(A) Cell cycle profiles of K562 cells before and after HS, as determined by FACS.

(B) GenomeDisco scores showing the reproducibility of the Hi-C libraries of G1/S phase cells.

(C) GenomeDisco scores showing the similarity among different conditions in G1/S phase cells.

(D) Fold changes of contact frequencies (observed / expected KR normalized) for Chr6 between

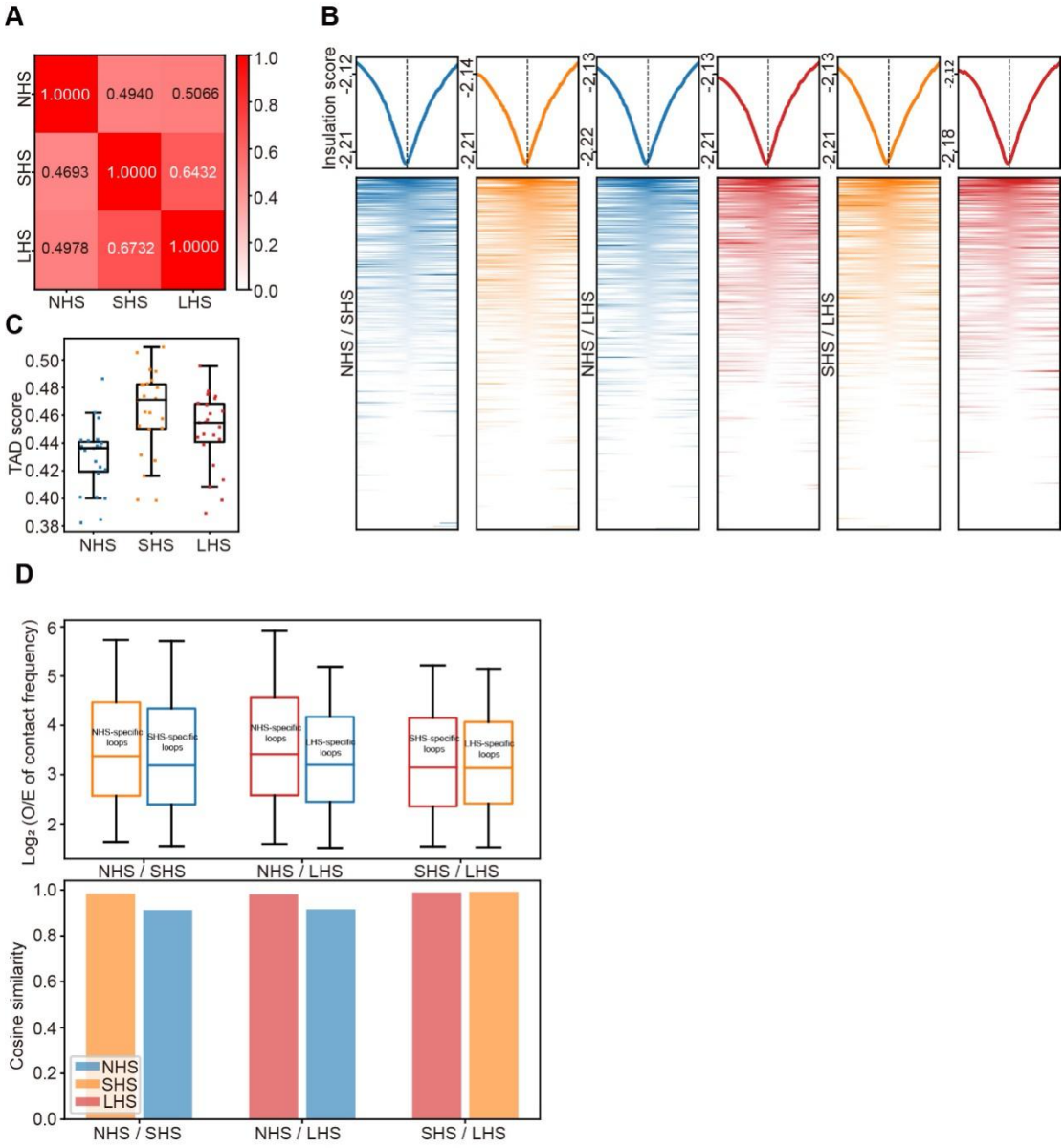
any pair of the three conditions for G1/S cells.

(E) Proportions of genome regions which switched compartments after SHS and LHS in G1/S phase cells.

(F) Proportions of genes (left) and dTREs (right) under compartment switches from NHS to SHS in G1/S cells.

(G) Insulation score differences in heat shock-regulated genes (left) and dTREs (right) in G1/S cells.

(H) Proportion of genes (left) and dTREs (right) under different regulation directions lying in anchors of condition- specific loops when comparing NHS with SHS in G1/S cells. UpLC: up regulated insignificantly, UpHC: up regulated significantly, UnReg: not significantly regulated, UnExp: not expressed both before and after heat shock, DownLC: down regulated insignificantly, DownHC: down regulated significantly.



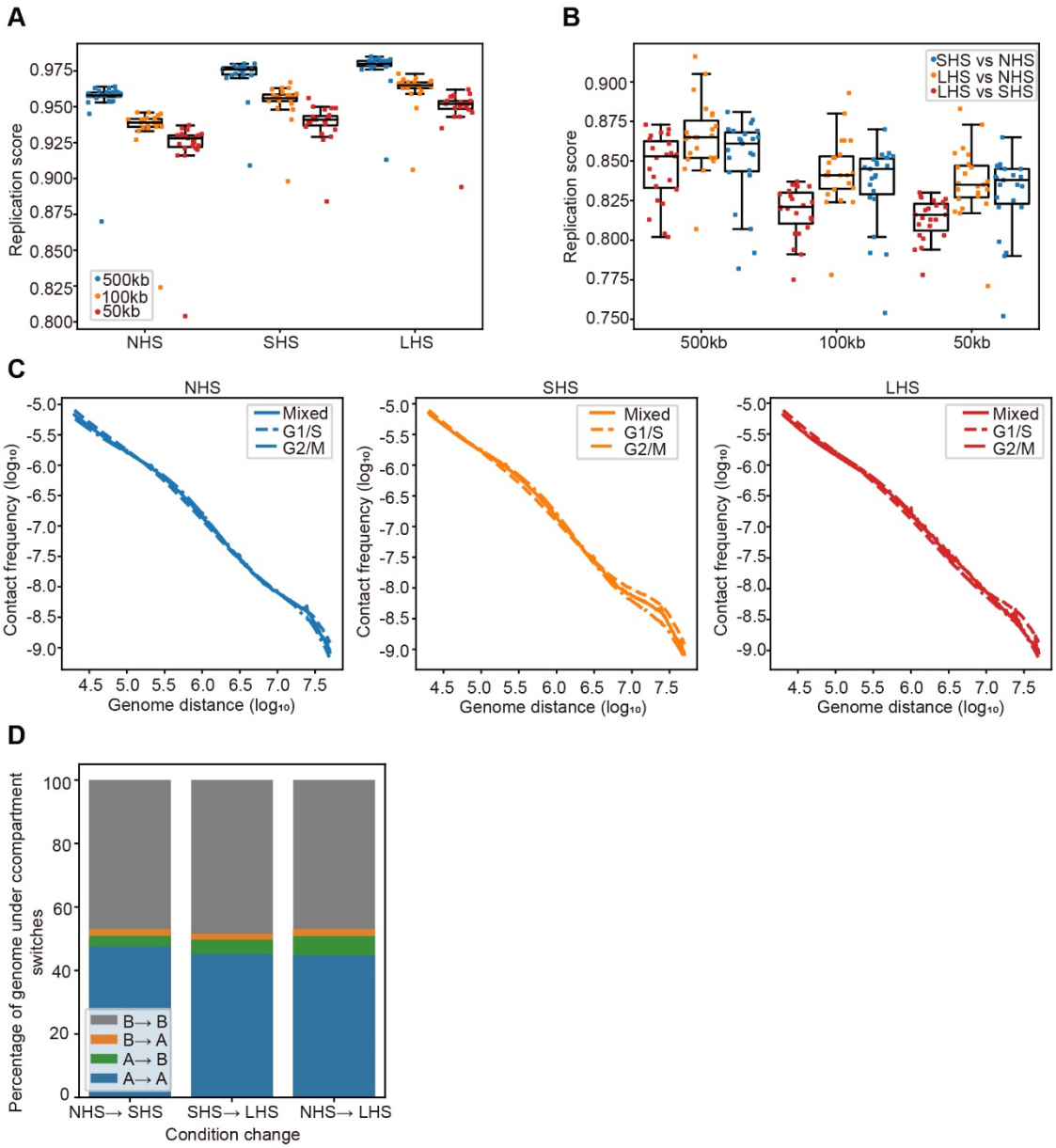
Supplemental Fig. S5. Changes of TADs and loops during HS in G1/S cells.

(A) Changes in TAD boundary during the process of HS in G1/S phase cells.

(B) The average and heatmap of insulation scores (±500kb) around condition-specific TAD boundaries in conditions not called. Signals were shown as paired comparison, and the conditions compared were marked on the left. Colors indicated the condition in which the insulation scores were calculated (NHS: blue, SHS: brown, LHS: red).

(C) TAD scores of G1/S cells at the three conditions.

(D) Comparison of contact frequencies of condition-specific loops in G1/S cells. When comparing the two conditions labeled in the X-axis, the upper boxplot showed the contact frequencies calculated in color-coded condition for specific loops called in the other condition. For example, the left orange boxplot showed the contact frequencies in SHS condition for NHS-specific loops. The lower panel showed the cosine similarity scores between the two labeled conditions in X-axis for loops specific to the color-coded condition. For example, the left orange bar showed the similarity between NHS and SHS for SHS-specific loops.



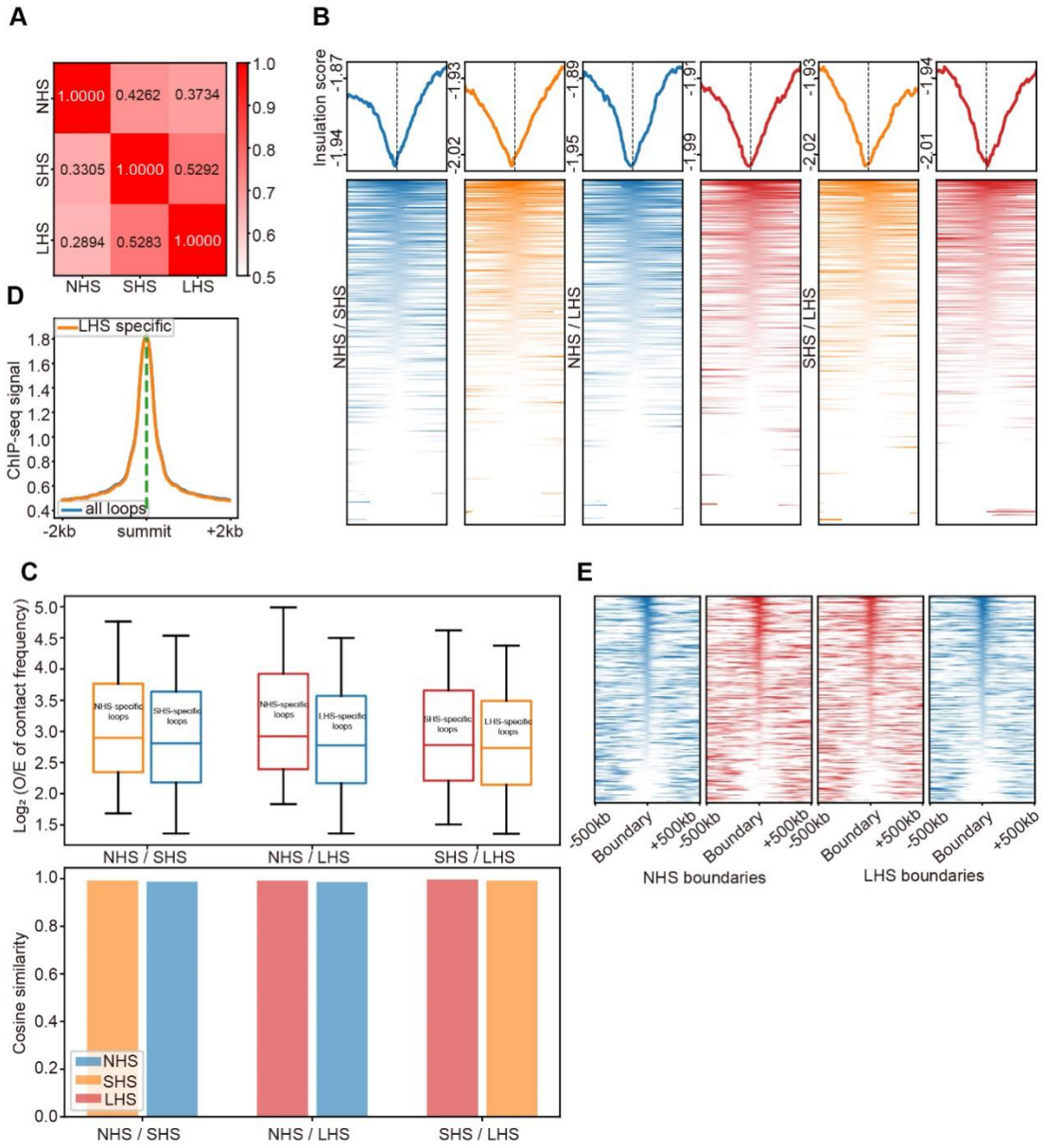
Supplemental Fig. S6. Chromatin conformation changes in G2/M phase cells during K562 HS.

(A) GenomeDisco scores between biological replicates, indicating the high quality of Hi-C libraries.

(B) GenomeDisco scores showing the similarity of Hi-C libraries of G2/M phase cells in different conditions.

(C) Comparison of contact frequency decay curves among G1/S, G2/M and mixed cells in the three conditions.

(D) The proportion of genome regions under compartment switches in G2/M phase cells.



Supplemental Fig. S7. Changes of TADs and loops during HS in G2/M cells.

(A) Changes in TAD boundary in G2/M phase cells during the process of HS. The values in row i and column j indicate the proportion of TAD boundaries in i which distanced no more than 1 bin with a boundary in condition j.

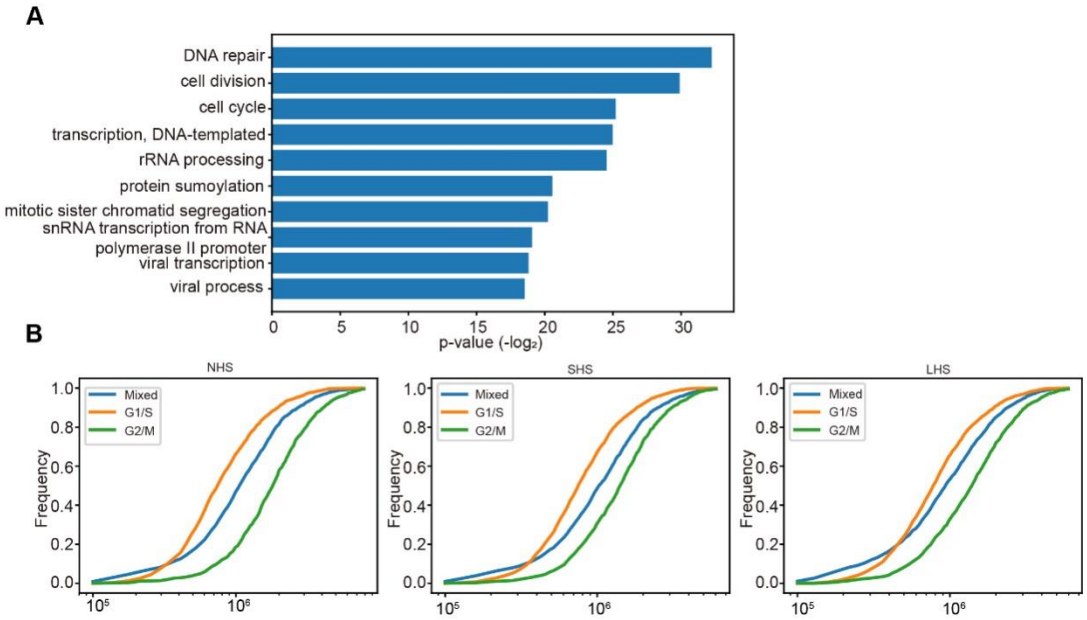
(B) The average and heatmap of insulation scores (±500kb) around condition-specific TAD boundaries in conditions not called. Signals were shown as paired comparison, and the conditions compared were marked on the left. Colors indicated the condition in which the insulation scores were calculated (NHS: blue, SHS: brown, LHS: red).

(C) Insulation score profiles around boundaries of loop domains in NHS and LHS for G2/M cells. The colors coded the conditions from which the profiles were calculated (blue: NHS, red: LHS).

(D) Comparison of contact frequencies of condition-specific loops in G2/M cells. When comparing the two conditions labeled in the X-axis, the upper boxplot showed the contact frequencies calculated in color-coded condition for specific loops called in the other condition. For example, the left orange boxplot showed the contact frequencies in SHS condition for NHS-specific loops.

The lower panel showed the cosine similarity scores between the two labeled conditions in X-axis for loops specific to the color-coded condition. For example, the left orange bar showed the similarity between NHS and SHS for SHS-specific loops.

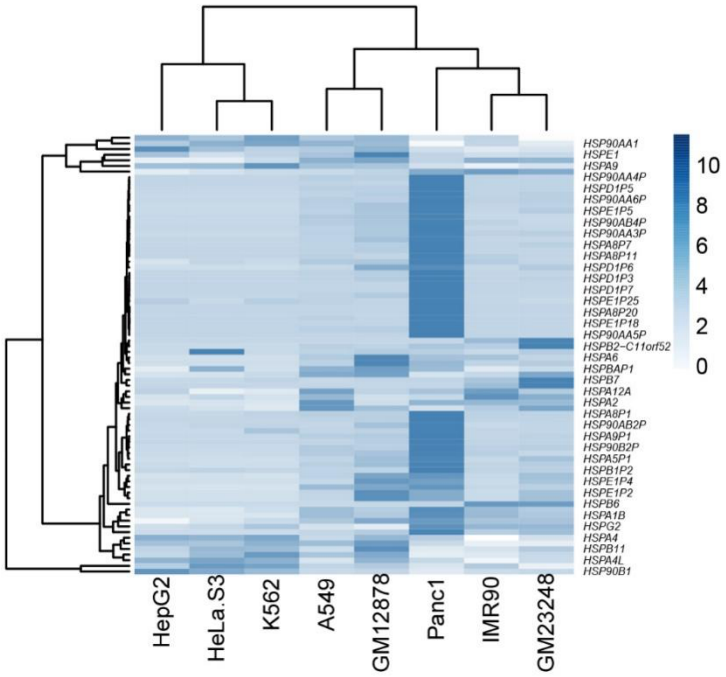
(E) CTCF ChIP-seq signals in anchors of LHS-specific loops and all combined loops in NHS condition for G2/M cells.



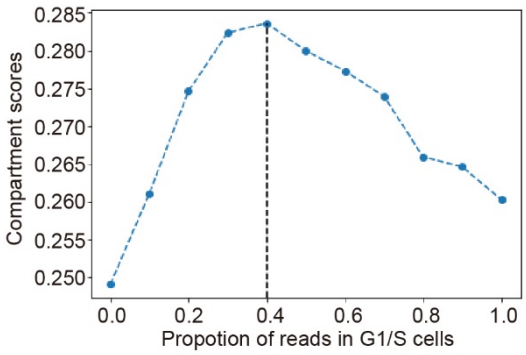
Supplemental Fig. S8. Chromatin conformation changes during the whole process of K562 heat shock was explained by cell cycle arrest.

(A) The biological processes which were mostly affected by the heat shock-induced transcription inhibition.

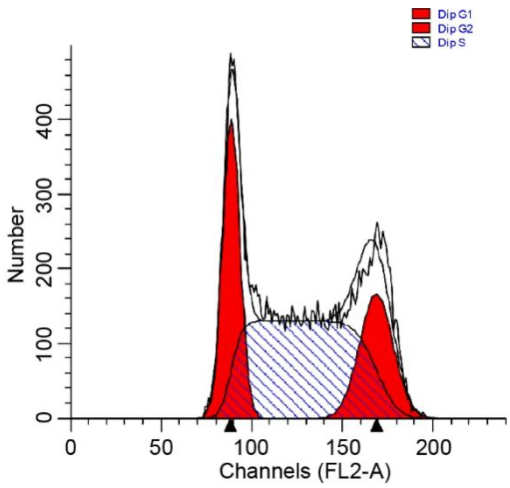
(B) Comparison of TAD lengths in mixed, G1/S and G2/M cells at the three conditions.



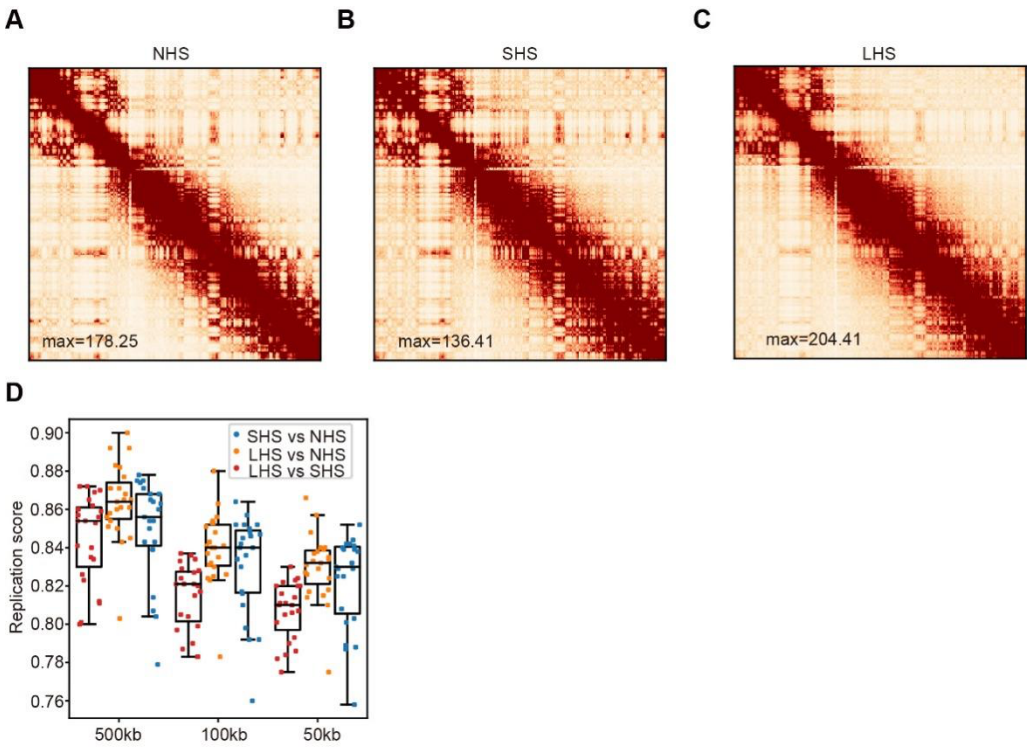
Supplemental Fig. S9. Gene expression profiles (square root of FPKM) of HSP family genes in 8 ENCODE cell lines. Only some gene names were listed owing to limited space.



Supplemental Fig. S10. Compartment scores of chr18 with mixed G1/S and G2/M cells.



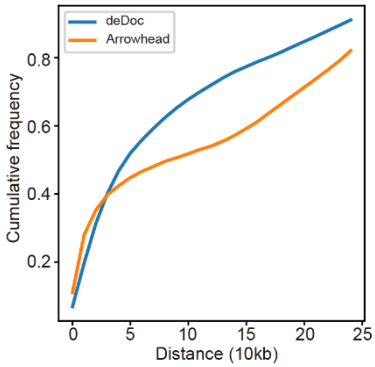
Supplemental Fig. S11. Cell stage distribution after heat shock of one hour in K562 cells. The percentage of G1, S and G2/M cells is 25.59%,54.22% and 20.18%, respectively.



Supplemental Fig. S12. Example of Hi-C results when aligned to GRCh38 genome.

(A-C) Contact maps of Chr6 in G2/M cells under NHS, SHS and LHS conditions, respectively.

(D) GenomeDisco scores between any pair of conditions in G2/M cells under different resolutions.



Supplemental Fig. S13. The cumulative distribution function of distances between called TAD boundaries and nearest local minima sites in the insulation score profiles for TAD called by deDoc and Arrowhead under NHS condition in mixed cells.

References

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159: 1665-1680.

Sanborn AL, Rao SS, Huang SC, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J et al. 2015. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A 112: E6456-6465.