Evolutionarily Conserved Principles Predict 3D Chromatin Organization

M. Jordan Rowley[1,3], Michael H. Nichols[1,3], Xiaowen Lyu[1], Masami Ando-Kuri[1], I. Sarahi M. Rivera[1], Karen Hermetz[1], Ping Wang[2], Yijun Ruan[2], and Victor G. Corces*[1]

SUPPLEMENTAL INFORMATION

Figures S1-S7

SUPPLEMENTAL FIGURES

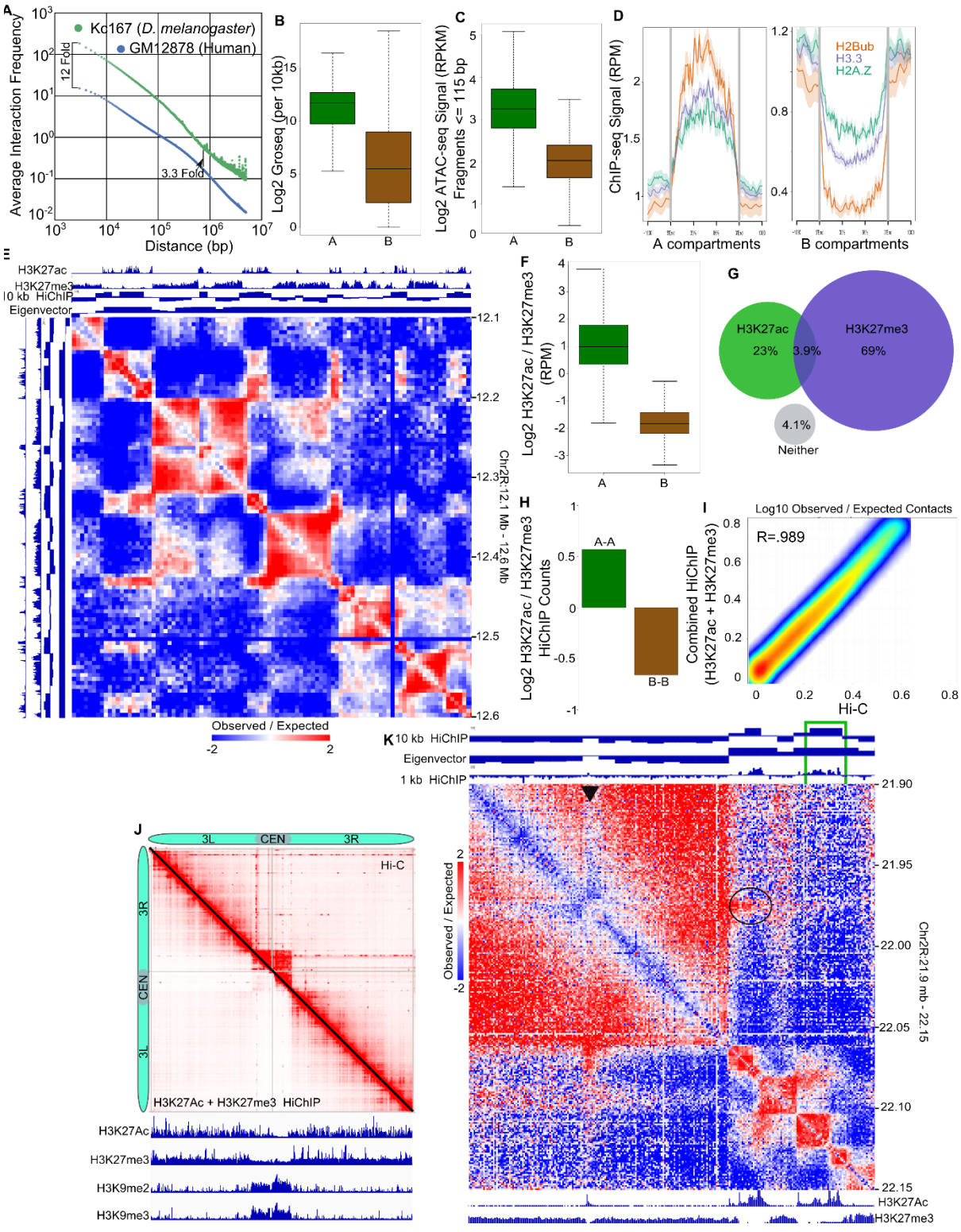

Figure S1 Related to Figure 1

A. Comparison of depth between the highest resolution *Drosophila* (green) and human (blue) Hi-C datasets at each distance.
B. A and B compartments have distinct GRO-seq signal. Log_2 GRO-seq signal in 10 kb windows on A (green) and B (brown) compartments. Boxes depict median and interquartile range.
C. A and B compartments have distinct transcription factor signal. Log_2 ATAC-seq signal corresponding to short protected fragments (<= 115 bp) in A (green) and B (brown) compartments. Boxes depict median and interquartile range.
D. Active and inactive chromatin correspond to A and B compartments. Average histone modification profiles over A and B compartments. Color coding of ChIP-seq for histone modifications/variants is indicated.
E. Hi-C PCA and eigenvector decomposition failed to identify small compartments. Distance normalized Hi-C (observed/expected). Tracks show H3K27ac ChIP-seq, H3K27me3 ChIP-seq, and the Hi-C eigenvector. Track with 10 kb HiChIP depicts the preference of each site to interact in H3K27ac or H3K27me3 HiChIP.
F. H3K27ac and H3K27me3 are good measures of A and B compartments. Log_2 H3K27ac/H3K27me3 ChIP-seq signal on A and B compartments. Boxes depict median and interquartile range.
G. Most of the *Drosophila* genome is contains either H3K27ac or H3K27me3. Percentage of 1 kb bins with enriched ChIP-seq signal for either H3K27ac (green), H3K27me3 (purple), or neither (grey).
H. H3K27ac and H3K27me3 capture A and B compartmental interactions, respectively. Interaction counts of inter-A (green) and inter-B (brown) compartments as log_2 ratio between HiChIP for H3K27ac and H3K27me3.
I. H3K27ac and H3K27me3 HiChIP combined recapitulates Hi-C interactions. Correlation between the merged HiChIP data and Hi-C. Distance normalized (observed/expected). Color range white to red indicates low to high density of data points.
J. H3K27ac and H3K27me3 HiChIP combined does not capture centromeric interactions. Hi-C compared to H3K27ac/H3K27me3 combined HiChIP across chromosome 3. ChIP-seq signal for H3K27ac, H3K27me3, H3K9me2, and H3K9me3 are shown below.
K. Compartmentalization occurs at 1 kb resolution. Distance normalized Hi-C (observed/expected). Tracks show H3K27ac ChIP-seq, H3K27me3 ChIP-seq (below), and the Hi-C eigenvector (above). The preference of each bin to interact in H3K27ac or H3K27me3 HiChIP computed at 10 kb and 1 kb resolution is shown. Arrows indicate compartmental switches discovered at 1 kb resolution at a small active site depleted for interactions with the surrounding inactive region. Circle indicates enriched interactions between the small active region and a nearby active region. Green box indicates compartmental switch missed by the eigenvector.

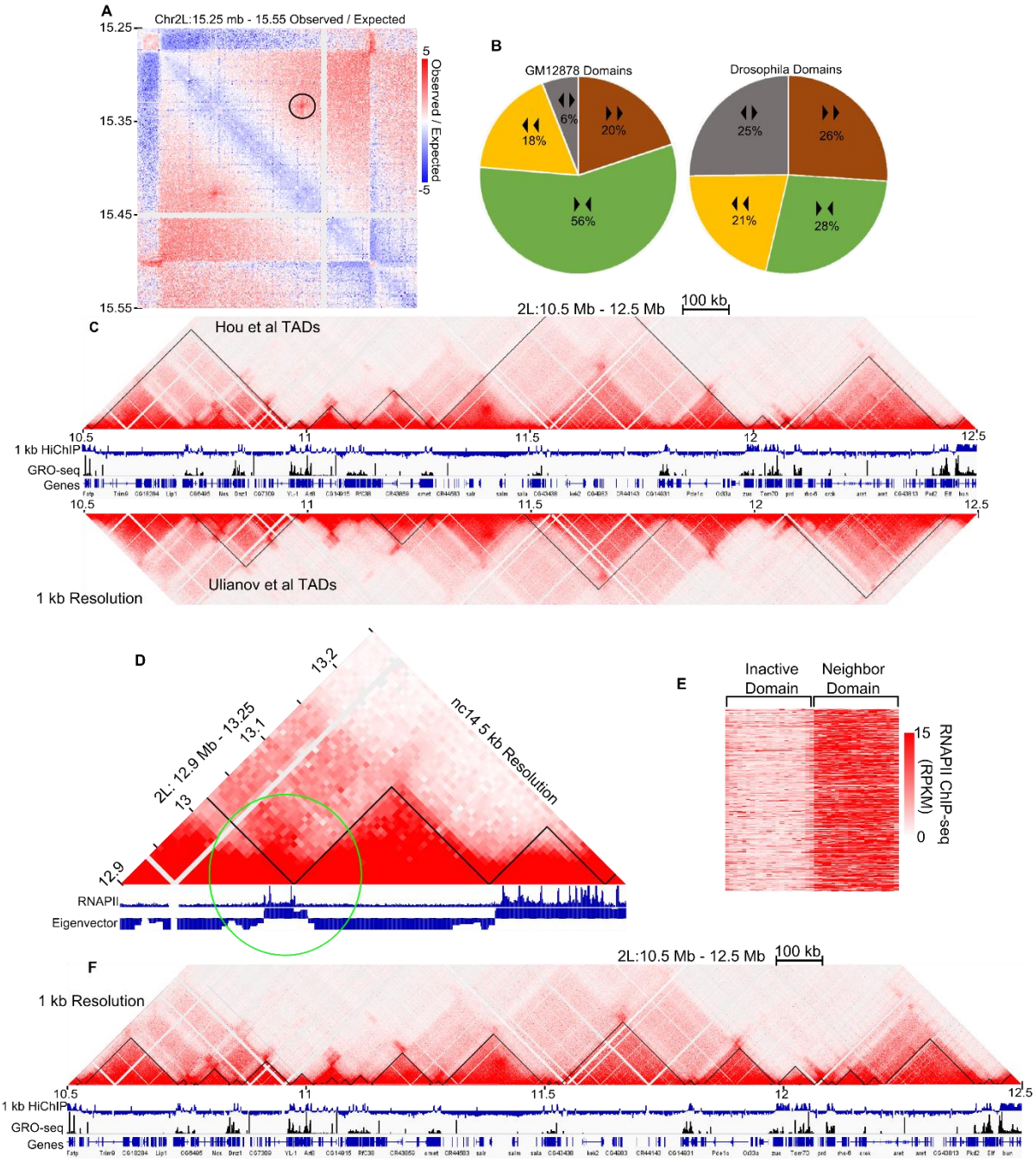

Figure S2 Related to Figure 2

A. *Drosophila* loops do not form domain corners. Distance normalized Hi-C depicting a loop (circled).

B. CTCF orientation does not determine domain corners in *Drosophila*. Percentage of domains with border associated CTCF motifs in each orientation for GM12878 cells and *Drosophila* Kc167 cells. Arrows indicate CTCF motif orientations.

C. Hi-C maps with previously identified TAD calls (black) (Hou et al., 2012) shown at 1 kb resolutions. GRO-seq, HiChIP determined compartmental associations, and gene

annotations are shown below. TAD calls from an independent source are also shown (bottom) (Ulianov et al., 2016).

D. Hi-C map of nc14 embryos with previously identified TAD calls (black) (Hug et al., 2017) shown at 5 kb resolution. RNAPII ChIP-seq signal and the compartmental associations (eigenvector) are shown below.

E. RNAPII is enriched throughout alternating domains. RNAPII ChIP-seq (RPKM) signal across neighboring domains.

F. Hi-C maps with compartmental domains (black) shown at 1 kb resolution. GRO-seq, HiChIP determined compartmental associations, and gene annotations are shown below.

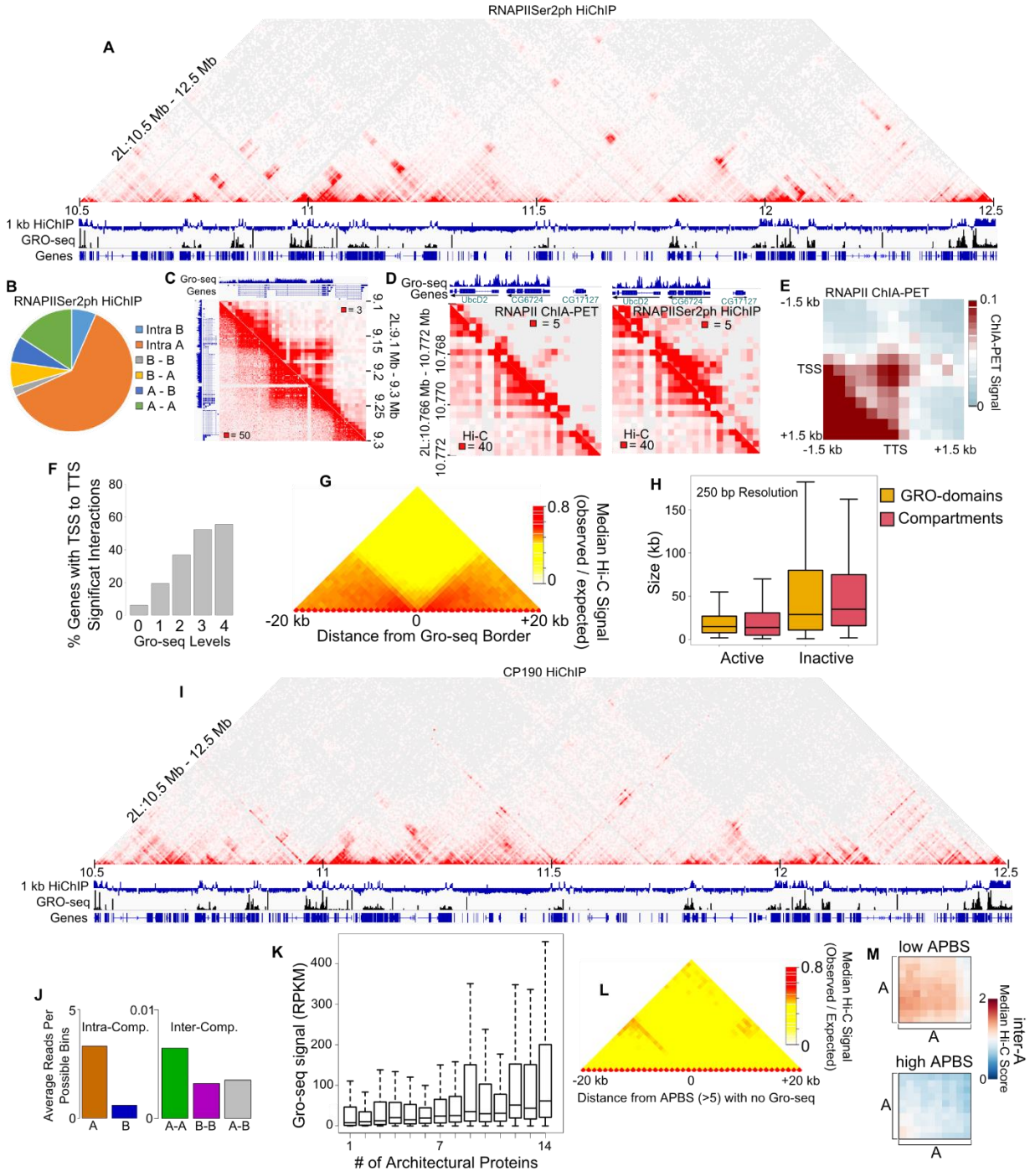

Figure S3 Related to Figure 2 and Figure 4

A. RNAPII HiChIP heatmap. GRO-seq, HiChIP determined compartmental associations, and gene annotations are shown below.

B. Active compartments are composed of elongating RNAPII. Categorization of contacts from HiChIP for RNAPIISer2ph and the enrichment within A and B compartments.

C. Individual genes can form mini contact domains. RNAPIISer2ph HiChIP (top right) compared to Hi-C signal (bottom left). GRO-seq and gene annotations are shown above and to the left.

D. Individual genes can form mini-domains. RNAPII ChIA-PET (left) and RNAPIISer2ph HiChIP (right) signal compared to Hi-C signal in 250 bp bins. GRO-seq and gene annotations are shown above.
E. RNAPII ChIA-PET displays gene-loops. Median metaplot of RNAPII ChIA-PET signal on genes.
F. Gene loops correspond to expression. Percentage of significant TSS-TTS interactions found by RNAPII ChIA-PET on genes without any GRO-seq signal (0) and in each quartile of expression (1-4).
G. GRO-seq borders correlate with Hi-C domain borders. Median distance normalized (observed/expected) Hi-C signal surrounding identified GRO-seq domain borders.
H. Sizes of *Drosophila* domains. Boxplot of the sizes of compartmental domains (orange) and domains called from GRO-seq data (pink). Boxes depict median and interquartile range.
I. CP190 HiChIP heatmap. GRO-seq, HiChIP determined compartmental associations, and gene annotations are shown below.
J. CP190 interactions are in active A compartments. Categorization of contacts from HiChIP for CP190 and the enrichment within A and B compartments.
K. Transcription corresponds to APBS occupancy. Boxplot of GRO-seq levels categorized by architectural protein occupancy within 250 bp of the TSS. Boxes depict median and interquartile range.
L. High occupancy APBSs do not demarcate Hi-C domain borders. Median distance normalized (observed/expected) Hi-C signal surrounding identified high occupancy (>= 5 protein occupancy) APBSs at least 20 kb away from highly expressed TSSs.
M. A compartments are further decayed by APBS occupancy. Median metaplot of A-A compartmental interactions for those separated by low (left) and high (right) APBSs.

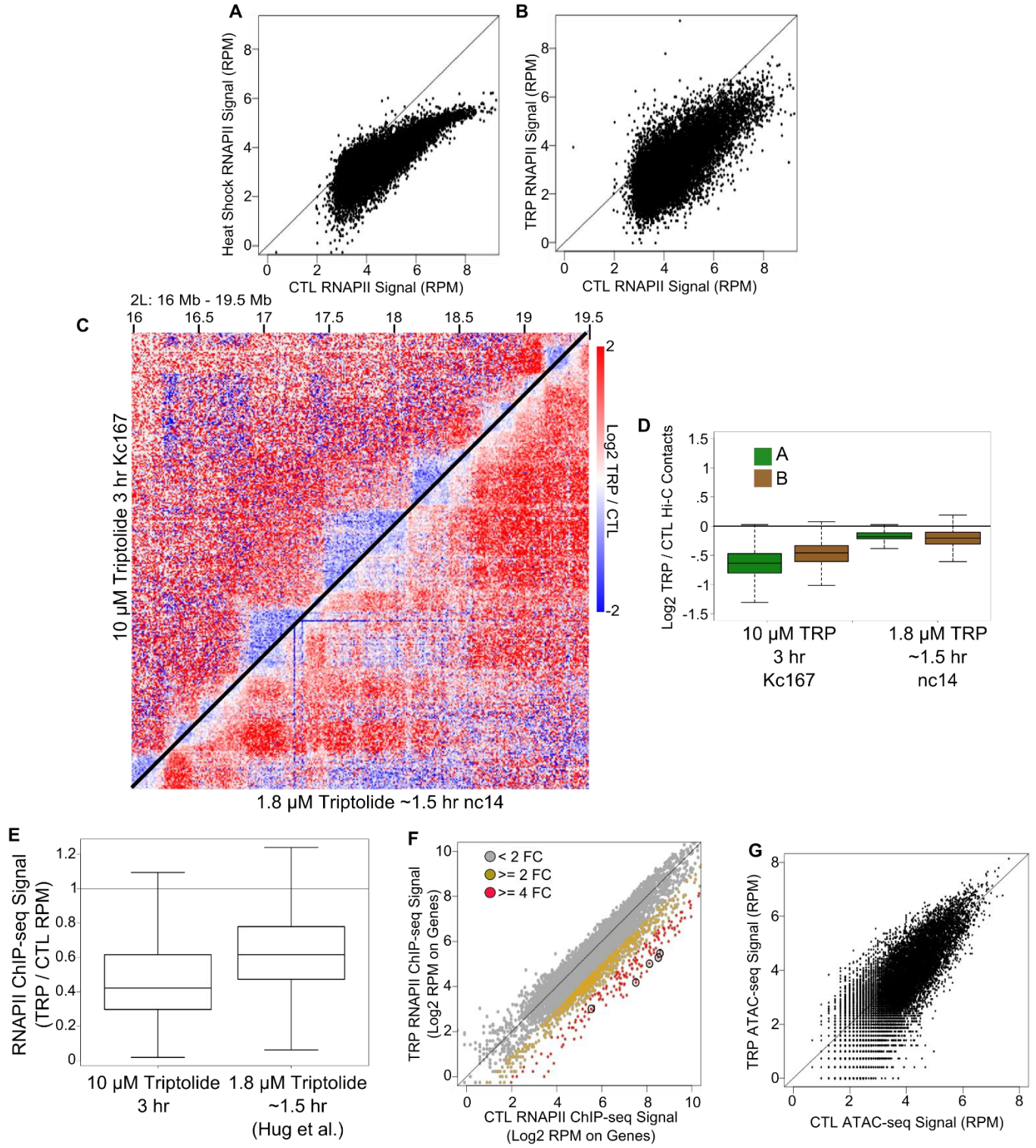

Figure S4 Related to Figure 3

A. Effect of heat shock on RNAPII. RNAPII ChIP-seq signal at peaks in control (CTL) vs heat shock.

B. Effect of triptolide on RNAPII. RNAPII ChIP-seq signal at peaks in control (CTL) vs triptolide treatment (TRP).
C. Comparison of high and low triptolide treatment on domain formation. Differential Hi-C signal (triptolide/control) for Kc167 cells with high treatment levels (top left) compared to nc14 with low treatment levels (Hug et al., 2017) (bottom right).
D. Treatment with higher concentration of triptolide for longer times correlates with a greater decrease in domain structure. Intra domain contact differences (triptolide/control) in A and B compartments in KC167 cells vs nc14 embryos(Hug et al., 2017) with different treatment conditions. Boxes depict median and interquartile range.
E. Higher triptolide treatment causes greater changes to RNAPII occupancy at peaks. RNAPII ChIP-seq differential signal (triptolide/control) in Kc167 cells vs nc14 embryos (Hug et al., 2017) with different treatment conditions. Boxes depict median and interquartile range.
F. Low triptolide treatment conditions have little effect on RNAPII within genes. Genes with RNAPII peaks in control were plotted for their total RNAPII signal in the control (CTL) vs triptolide (TRP). Data from Hug et al. were used (Hug et al., 2017). Line indicates slope of 1. Grey, yellow, and red dots represent genes with less than 2, 2-4, and greater than 4-fold decrease in RNAPII ChIP-seq signal respectively. Black circles indicate genes tested by RT-PCR in Hug et al.
G. Effect of triptolide on chromatin accessibility. ATAC-seq signal (fragments <= 115 bp) on peaks in in control (CTL) vs triptolide treatment (TRP).

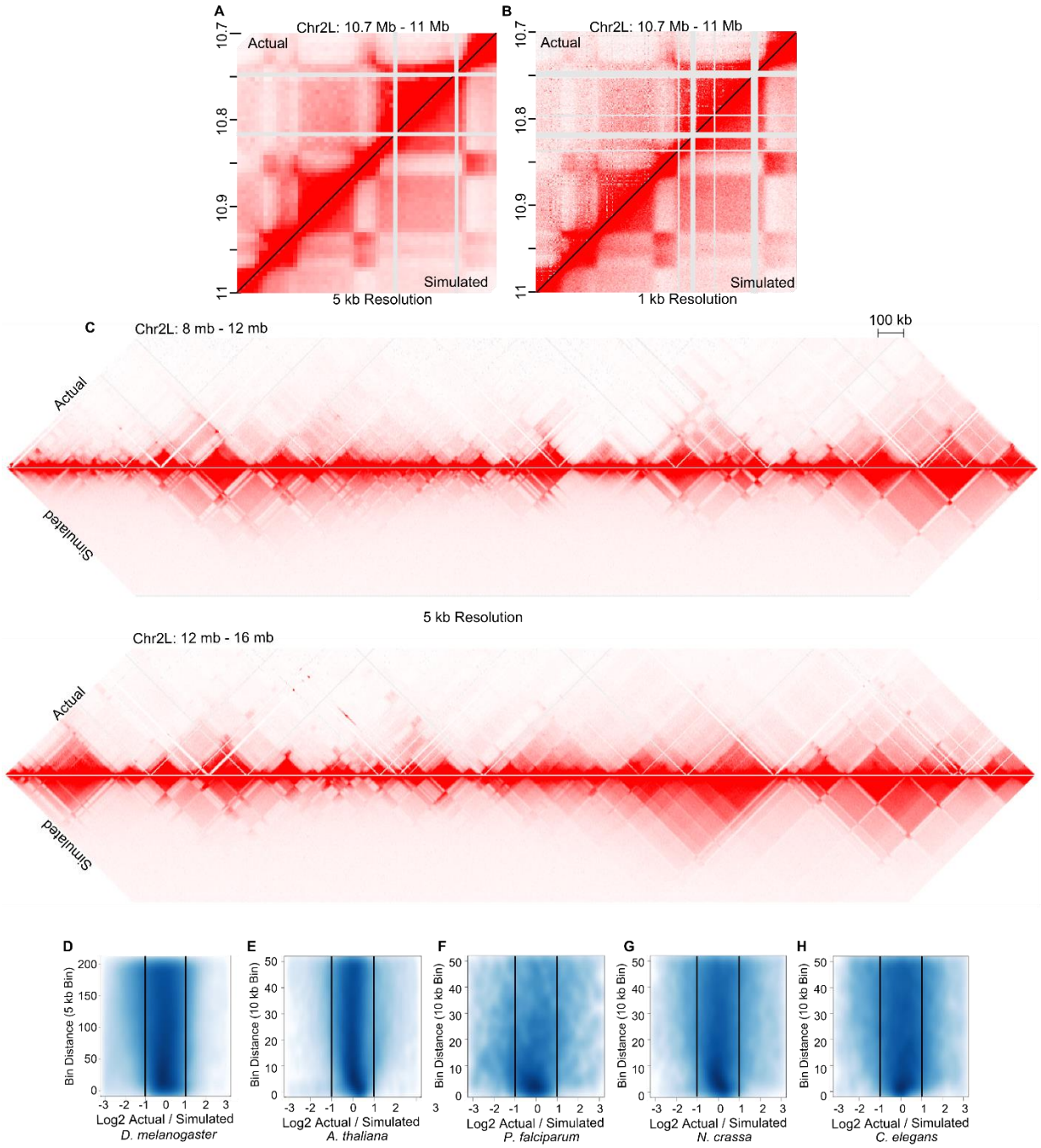

Figure S5 Related to Figure 5

A. GRO-seq and APBS based simulated contact maps precisely recapitulate domains and compartments at 5 kb resolution. Contact heatmap at 5 kb resolution using actual Hi-C data (top) and simulated data based on GRO-seq and APBS occupancy (bottom). 300 kb of chr2L are shown.

B. GRO-seq and APBS based simulated contact maps precisely recapitulate domains and compartments at 1 kb resolution. Contact heatmap at 1 kb resolution using actual Hi-C data

(top) and simulated data based on GRO-seq and APBS occupancy (bottom). 300 kb of chr2L are shown.

C. GRO-seq and APBS based simulated contact maps recapitulate domains and compartments. Contact heatmaps at 5 kb resolution using actual Hi-C data (top) and simulated data based on GRO-seq and APBS occupancy (bottom).

D-H. Actual Hi-C compared to simulated data. Density scatter plot showing the ratio of actual Hi-C reads over the simulated reads (x-axis) for each genomic distance (y-axis) for *D. melanogaster* (D), *A. thaliana* (E), *P. falciparum* (F), *N. crassa* (G), *C. elegans* (H). Reads were normalized by distance decay values. Vertical lines indicate a 2-fold change.

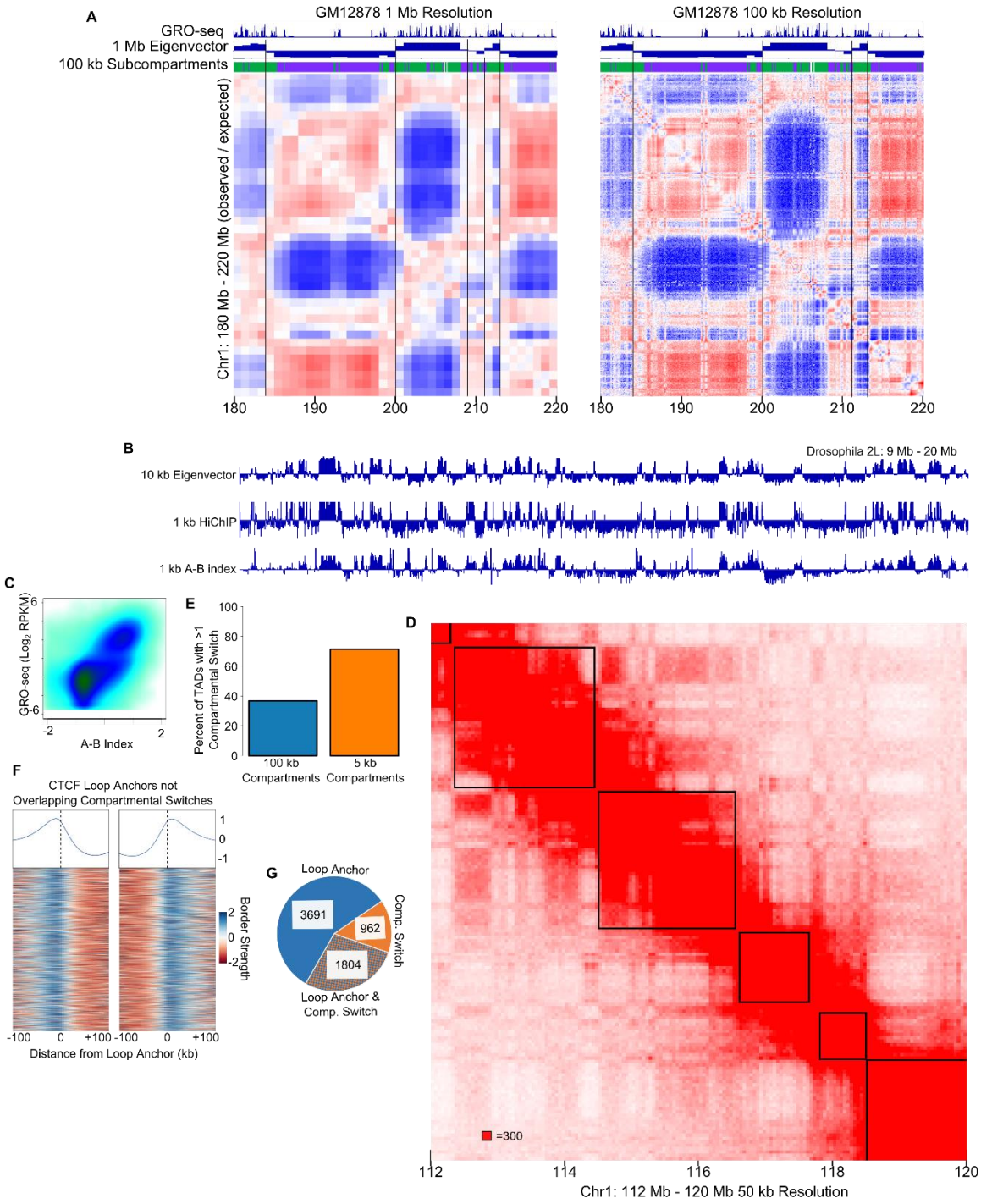

Figure S6 Related to Figure 6

A. Large compartments are a result of read binning. Distance normalized Hi-C (observed/expected) in GM12878 cells at 1 Mb (left) and 100 kb (right) resolution. Tracks with 1 Mb eigenvector, 100 kb compartments, and GRO-seq are shown above.
B. Validation of the A/B index compartmental refinement method. *Drosophila* tracks comparing the A-B index to the eigenvector or 1 kb HiChIP compartment calls.

C. Transcriptional activity corresponds to fine-scale compartments. Log_2 GRO-seq compared to A-B index. Color gradient indicates density of points.
D. TAD calls at saturation. TAD calls in GM12878 cells at 50 kb resolution.
E. TADs are comprised of compartmental interactions. Percentage of TADs with more than 1 compartment switch within the domain. Previously called (at 100 kb resolution - blue) and newly called (at 5 kb resolution – orange) compartments are shown.
F. CTCF loops correlate with borders. Heatmap displaying the border strength for 100 kb to either side of left and right CTCF loop anchors. Average profiles are shown above.
G. TAD borders correspond to loops and compartments. Number of TAD borders coinciding with loops, compartmental switches, or both. $p < .001$ for loop anchors and compartments via random permutation test.

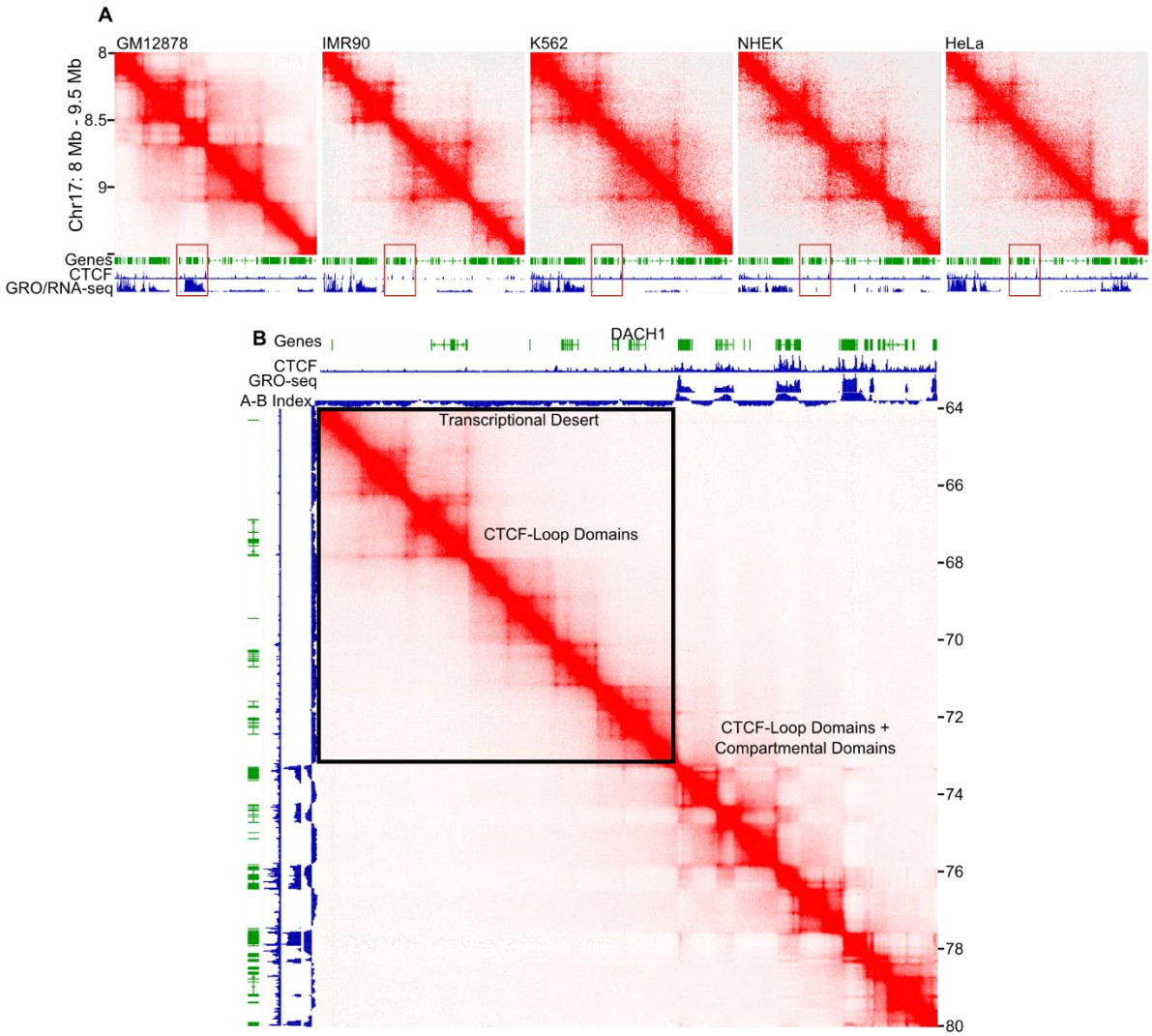

Figure S7 Related to Figure 7

A. Transcriptional activity correlates with the formation of domains. Hi-C heatmaps for GM12878, IMR90, K562, NHEK, and HeLa cells. Tracks comparing GRO-seq/RNA-seq and CTCF occupancy between the different cell lines are shown below. Red rectangle indicates differentially expressed region.

B. Compartmental domains underlie CTCF loops. Hi-C heatmap for GM12878 cells displaying a large region with CTCF loops and no compartmental domains (transcriptional desert), as well as a region with compartmental domains and CTCF loops. Gene annotations, CTCF ChIP-seq, GRO-seq, and fine-scale compartment identification tracks are shown above and to the left.

	H3K27ac Rep1 HiChIP	H3K27ac Rep2 HiChIP
Sequenced Read Pairs	100,225,894	22,784,945
Normal Paired	80,184,199 (80.00%)	18,663,253 (81.91%)
Chimeric Paired	1,066 (0.00%)	119 (0.00%)
Chimeric Ambiguous	2,002 (0.00%)	390 (0.00%)
Unmapped	20,038,627 (19.99%)	4,121,183 (18.09%)
Ligation Motif Present	31,666,024 (31.59%)	8,548,749 (37.52%)
Alignable (Normal+Chimeric Paired)	80,185,265 (80.00%)	18,663,372 (81.91%)
Unique Reads	68,648,039 (68.49%)	17,945,684 (78.76%)
PCR Duplicates	11,536,419 (11.51%)	716,899 (3.15%)
Optical Duplicates	807 (0.00%)	789 (0.00%)
Library Complexity Estimate	251,239,520	236,654,145
Intra-fragment Reads	889,140 (0.89% / 1.30%)	1,627,771 (7.14% / 9.07%)
Below MAPQ Threshold	18,799,325 (18.76% / 27.39%)	4,793,044 (21.04% / 26.71%)
Hi-C Contacts	48,959,574 (48.85% / 71.32%)	11,524,869 (50.58% / 64.22%)
Ligation Motif Present	12,269,091 (12.24% / 17.87%)	3,604,694 (15.82% / 20.09%)
3' Bias (Long Range)	85% - 15%	78% - 22%
Pair Type %(L-I-O-R)	25% - 25% - 25% - 25%	25% - 25% - 25% - 25%
Inter-chromosomal	2,632,278 (2.63% / 3.83%)	786,657 (3.45% / 4.38%)
Intra-chromosomal	46,327,296 (46.22% / 67.49%)	10,738,212 (47.13% / 59.84%)
Short Range (<20Kb)	23,349,446 (23.30% / 34.01%)	5,133,320 (22.53% / 28.60%)
Long Range (>20Kb)	22,975,053 (22.92% / 33.47%)	5,604,850 (24.60% / 31.23%)
Combined Contacts: 60,484,443		

Table S1 Related to Figure 1
HiChIP and ChIA-PET mapping statistics performed in Kc167 cells to the dm6 genome for H3K27ac.

	H3K27me3 Rep1 HiChIP	H3K27me3 Rep2 HiChIP	H3K27me3 Rep3 HiChIP
Sequenced Read Pairs	86,473,347	22,785,730	73,918,791
Normal Paired	68,220,334 (78.89%)	19,171,480 (84.14%)	54,625,690 (73.90%)
Chimeric Paired	1,421 (0.00%)	259 (0.00%)	302 (0.00%)
Chimeric Ambiguous	2,149 (0.00%)	640 (0.00%)	1,308 (0.00%)
Unmapped	18,249,443 (21.10%)	3,613,351 (15.86%)	19,291,491 (26.10%)
Ligation Motif Present	27,127,062 (31.37%)	5,362,368 (23.53%)	31,273,691 (42.31%)
Alignable (Normal+Chimeric Paired)	68,221,755 (78.89%)	19,171,739 (84.14%)	54,625,992 (73.90%)
Unique Reads	18,003,439 (20.82%)	11,579,237 (50.82%)	46,920,236 (63.48%)
PCR Duplicates	50,217,579 (58.07%)	7,591,992 (33.32%)	7,704,177 (10.42%)
Optical Duplicates	737 (0.00%)	510 (0.00%)	1,579 (0.00%)
Library Complexity Estimate	18,462,085	17,270,710	174,978,194
Intra-fragment Reads	304,152 (0.35% / 1.69%)	340,787 (1.50% / 2.94%)	1,328,032 (1.80% / 2.83%)
Below MAPQ Threshold	6,606,156 (7.64% / 36.69%)	4,286,041 (18.81% / 37.01%)	17,587,159 (23.79% / 37.48%)
Contacts	11,093,131 (12.83% / 61.62%)	6,950,409 (30.51% / 60.04%)	28,005,045 (37.89% / 59.69%)
Ligation Motif Present	2,951,454 (3.41% / 16.39%)	1,306,510 (5.73% / 11.28%)	9,483,361 (12.83% / 20.21%)
3' Bias (Long Range)	84% - 16%	74% - 26%	78% - 22%
Pair Type %(L-I-O-R)	25% - 25% - 25% - 25%	25% - 25% - 25% - 25%	25% - 25% - 25% - 25%
Inter-chromosomal	975,417 (1.13% / 5.42%)	534,655 (2.35% / 4.62%)	1,839,742 (2.49% / 3.92%)
Intra-chromosomal	10,117,714 (11.70% / 56.20 %)	6,417,754 (28.17% / 55.42%)	26,165,303 (35.40% / 55.77%)
Short Range (<20Kb)	3,935,392 (4.55% / 21.86%)	3,103,851 (13.62% / 26.81%)	11,830,360 (16.00% / 25.21%)
Long Range (>20Kb)	6,180,938 (7.15% / 34.33%)	3,313,884 (14.54% / 28.62%)	14,334,925 (19.39% / 30.55%)
Combined Contacts: 46,048,585			

Table S2 Related to Figure 1

HiChIP and ChIA-PET mapping statistics performed in Kc167 cells to the dm6 genome for H3K27me3.

Experiment description	RNAPIISer2ph HiChIP Rep 1	RNAPIISer2ph HiChIP Rep2	RNAPII ChIA-PET	CP190 HiChIP
Sequenced Read Pairs	34,167,551	73,194,819	33,708,212	112,982,203
Normal Paired	30,167,498 (88.29%)	62,999,182 (86.07%)	21,050,572 (62.45%)	89,231,912 (78.98%)
Chimeric Paired	2,513 (0.01%)	425 (0.00%)	3,821,972 (11.34%)	9,957 (0.01%)
Chimeric Ambiguous	901 (0.00%)	1,219 (0.00%)	7,570,272 (22.46%)	2,803 (0.00%)
Unmapped	3,996,639 (11.70%)	10,193,993 (13.93%)	1,265,396 (3.75%)	23,737,531 (21.01%)
Ligation Motif Present	5,724,083 (16.75%)	17,071,721 (23.32%)	88,205 (0.26%)	23,426,133 (20.73%)
Alignable (Normal+Chimeric Paired)	30,170,011 (88.30%)	62,999,607 (86.07%)	24,872,544 (73.79%)	89,241,869 (78.99%)
Unique Reads	23,764,884 (69.55%)	57,498,996 (78.56%)	14,769,195 (43.81%)	62,817,133 (55.60%)
PCR Duplicates	6,403,207 (18.74%)	5,498,867 (7.51%)	10,098,920 (29.96%)	26,424,013 (23.39%)
Optical Duplicates	1,920 (0.01%)	1,744 (0.00%)	4,429 (0.01%)	723 (0.00%)
Library Complexity Estimate	60,608,219	339,548,156	21,598,913	119,190,245
Intra-fragment Reads	1,320,369 (3.86% / 5.56%)	3,619,547 (4.95% / 6.29%)	7,539,437 (22.37% / 51.05%)	9,756,188 (8.64% / 15.53%)
Below MAPQ Threshold	6,546,934 (19.16% / 27.55%)	13,203,639 (18.04% / 22.96%)	1,431,119 (4.25% / 9.69%)	15,674,333 (13.87% / 24.95%)
Contacts	15,897,581 (46.53% / 66.90%)	40,675,810 (55.57% / 70.74%)	5,798,639 (17.20% / 39.26%)	37,386,612 (33.09% / 59.52%)
Ligation Motif Present	2,148,740 (6.29% / 9.04%)	7,427,401 (10.15% / 12.92%)	35,778 (0.11% / 0.24%)	6,873,662 (6.08% / 10.94%)
3' Bias (Long Range)	80% - 20%	73% - 27%	51% - 49%	79% - 21%
Pair Type %(L-I-O-R)	25% - 25% - 25% - 25%	25% - 25% - 25% - 25%	25% - 25% - 25% - 25%	25% - 25% - 25% - 25%
Inter-chromosomal	1,027,932 (3.01% / 4.33%)	1,332,949 (1.82% / 2.32%)	281,654 (0.84% / 1.91%)	950,358 (0.84% / 1.51%)
Intra-chromosomal	14,869,649 (43.52% / 62.57%)	39,342,861 (53.75% / 68.42%)	5,516,985 (16.37% / 37.35%)	36,436,254 (32.25% / 58.00%)
Short Range (<20Kb)	8,921,240 (26.11% / 37.54%)	24,659,290 (33.69% / 42.89%)	4,893,337 (14.52% / 33.13%)	26,353,410 (23.33% / 41.95%)

Long Range (>20Kb)	5,935,988 (17.37% / 24.98%)	14,683,333 (20.06% / 25.54%)	623,626 (1.85% / 4.22%)	10,074,135 (8.92% / 16.04%)
	Combined Contacts: 56,573,391			

Table S3 Related to Figure 2 and Figure 4
HiChIP and ChIA-PET mapping statistics performed in Kc167 cells to the dm6 genome for RNAPIISer2ph, CP190, and RNAPII.