SI APPENDIX

Polycomb-Mediated Chromatin Loops Revealed by a Sub-Kilobase Resolution Chromatin Interaction Map

Kyle P. Eagen\textsuperscript{a,1,2}, Erez Lieberman Aiden\textsuperscript{b,c,d,e}, Roger D. Kornberg\textsuperscript{a,2}

\textsuperscript{a}Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA
\textsuperscript{b}The Center for Genome Architecture, Baylor College of Medicine, Houston, TX 77030, USA
\textsuperscript{c}Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
\textsuperscript{d}Center for Theoretical Biological Physics, Rice University, Houston, TX 77005, USA
\textsuperscript{e}Department of Computer Science, Rice University, Houston, TX 77005, USA
\textsuperscript{1}Present address: Department of Biochemistry and Molecular Genetics, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA
\textsuperscript{2}To whom correspondence should be addressed. Email: \texttt{eagen@northwestern.edu} and \texttt{kornberg@stanford.edu}

The Hi-C sequence data reported in this paper have been deposited at the GEO database, \url{www.ncbi.nlm.nih.gov/geo} (accession no. GSE89112).
SUPPLEMENTARY MATERIALS AND METHODS

Cell Culture

Kc167 cells were obtained from the Drosophila Genomics Resource Center (Bloomington, IN). Cells were grown in CCM3 media (HyClone) at 25°C.

Hi-C Library Preparation

Hi-C libraries were prepared using a modification of the previously described tethered conformation capture (TCC) (1, 2) protocol. 1 billion Kc167 cells at a density of 4-6 x 10^6 cells/mL were washed with 20 mL CMM3 media and then cross-linked with 1% EM-grade paraformaldehyde in 450 mL CCM3 media for 10 minutes at room temperature while mixing. Paraformaldehyde was quenched by adding 29.8 mL 2.5 M glycine (150 mM final) and incubating for 5 minutes at room temperature while mixing. Cells were washed with 10 mL ice-cold PBS, pelleted, flash frozen in liquid nitrogen and stored at -80°C.

Cells were thawed, resuspended in 10 mL lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% Igepal CA-630, 1 mM PMSF, 2 mM bezamidine, 2 μM pepstatin A, 0.6 μM leupeptin), and then incubated on ice for 15 minutes. Cells were transferred to a 15 mL Dounce homogenizer and lysed by applying 15 strokes of pestle B, cooled on ice for 1 minute, followed by applying another 15 strokes of pestle B. The lysate was centrifuged at 2,500 x g for 5 minutes, the pellet washed with 5 mL lysis buffer, centrifuged again at 2,500 x g for 5 minutes, the pellet resuspended in 4.6 mL of wash buffer 1 (50 mM Bis-
Tris-HCl pH 6.0, 100 mM NaCl, 10 mM MgCl₂, 0.1% SDS), incubated at 65°C for 10 minutes, and then immediately cooled on ice.

1.4 mL 25 mM EZ-link iodoacetyl-PEG2-Biotin (IPB; in wash buffer 1; Thermo Scientific) was added to the sample and incubated for 1 hour at room temperature while rotating. 600 μL of 10% Triton X-100 was added, gently mixed, then 160 μL of 50 U/μL DpnII (NEB) was added and the sample incubated at 37°C for 2 hours while rotating. 1.298 mL 10% SDS was added, incubated at 65°C for 30 minutes, and then immediately cooled on ice. The sample was added to a 20 kD MWCO, 3-12 mL Slide-A-Lyzer Dialysis Cassette (Thermo Scientific) and dialyzed at room temperature against 4 L of dialysis buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). After 2 hours the buffer was replaced with 4 L of fresh dialysis buffer and dialysis continued overnight.

800 μL of Dynabeads MyOne Streptavidin T1 beads (Life Technologies) were washed three times with 2 mL PBST (PBS + 0.01% Tween 20) and then resuspended in 1 mL PBST. The dialyzed sample was rapidly added to the beads and incubated at room temperature for 30 minutes while rotating. 50 μL of 25 mM neutralized IPB (treated with 10-fold excess β-mercaptoethanol) was added to the sample and then incubated at room temperature for 15 minutes while rotating.

Beads were washed once with 3 mL PBST, twice with 3 mL wash buffer 2 (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 0.4% Triton X-100), resuspended in 1 mL of wash buffer 2,
and divided across 5 equal volume aliquots. 190 μL of a fill-in master mix (654 μL water, 11 μL 1 M MgCl₂, 110 μL NEBuffer 2, 7.7 μL 10 mM dTTP, 7.7 μL 10 mM dCTP, 7.7 μL 10 mM dGTPαS, 192.5 μL 0.4 mM Biotin-14-dATP, 44 μL 10% Triton X-100) was added to each aliquot followed by 10 μL of 5 U/μL DNA Polymerase I, Large (Klenow) Fragment (NEB) and then incubated at 37° C for 75 minutes while rotating. The reaction was stopped with 10 μL 0.5 M EDTA.

Each aliquot of beads was washed twice with 500 μL wash buffer 3 (50 mM Tris-HCl pH 7.4, 0.4% Triton X-100, 0.1 mM EDTA) and then resuspended in 500 μL of wash buffer 3. 8.99 mL of a ligation master mix (37.972 mL water, 5.179 mL 10% Triton X-100, 5.179 mL 500 mM Tris-HCl pH 7.5, 100 mM MgCl₂, 556 μL 10 mg/mL BSA, 556 μL 100 mM ATP, 556 μL 1 M DTT) was added to each aliquot followed by 10 μL 2,000 U/μL T4 DNA Ligase (NEB) and then incubated at room temperature for 4 hours while rotating. The reaction was stopped with 400 μL 0.5 M EDTA.

Beads were collected with a magnet, the supernatant discarded, and the beads resuspended in 300 μL of extraction buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.2% SDS). 25 μL 800 U/mL proteinase K (NEB) was added to each aliquot and each aliquot was incubated at 65° C overnight.
Beads were collected with a magnet and the supernatant transferred to a fresh tube. The DNA was precipitated with ethanol, washed twice with 70% ethanol, and all aliquots combined and resuspended in a total volume of 75 μL 10 mM Tris-HCl pH 8.0. The sample was incubated at 42° C for 15 minutes followed by addition of 5 μL 1 mg/mL RNase A and incubation at 37° C for 30 minutes. DNA concentration was determined with a Qubit dsDNA HS Assay.

7 μL water, 10 μL 10x NEBuffer 1, 1 μL 10 mg/mL BSA, and 3 μL 100 U/μL exonuclease III was added to 79 μL of DNA (no more than 10 μg DNA per reaction, if more than 10 μg multiple reactions were performed in parallel), and incubated at 37° C for 1 hour. The reaction was stopped by adding 2 μL 0.5 M EDTA, 2 μL 5 M NaCl, and incubating at 70° C for 20 minutes. Total sample volume was adjusted to 130 μL.

DNA was sheared to 500 bp with a Covaris S2 at duty cycle 10%, intensity 4, 200 cycles/burst for 55 seconds. 125 μL of sample was transferred to a fresh tube and the DNA was size-selected by first adding 68.8 μL (0.55x volumes) of SPIRselect beads (Beckman), vortexed to mix, and incubated at room temperature for 5 minutes. Beads were collected with a magnet, the supernatant transferred to a fresh tube. 25 μL of SPIRselect beads were added to the supernatant, vortexed to mix, and incubated at room temperature for 5 minutes. Beads were collected with a magnet, the supernatant discarded, and, while still on the magnet, the beads were washed twice with 200 μL.
85% ethanol and then air dried. DNA was eluted by resuspending the beads in 53 μL of 10 mM Tris-HCl pH 8.0 and incubated at room temperature for 5 minutes. Beads were collected with a magnet, and 51 μL of the eluate was transferred to a fresh tube. DNA concentration was determined with a Qubit dsDNA HS Assay.

DNA was initially prepared for high-throughput sequencing following the directions for “NEBNext End Prep” and “Adaptor Ligation” in the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB) (if more than 1 μg of DNA, multiple reactions were performed in parallel). 50 μL of Dynabeads MyOne Streptavidin C1 beads (Life Technologies) were washed twice with 100 μL 1x B&W buffer + 0.05% Tween 20 (2x Bind & Wash (B&W) buffer: 10 mM Tris-HCl pH 7.5, 2 M NaCl, 1 mM EDTA) and then resuspended in 86.5 μL of 2x B&W buffer. Adaptor-ligated DNA was added to the beads and incubated at room temperature for 30 minutes while rotating. Beads were collected with a magnet, the supernatant discarded, the beads washed twice with 100 μL 1x B&W buffer + 0.05% Tween 20, and then twice with 100 μL 10 mM Tris-HCl pH 8.0, 1 mM EDTA.

DNA was eluted by resuspending the beads in 5 μL of 95% freshly deionized formamide, 10 mM EDTA pH 8.0, and incubating at 90° C for 10 minutes. The beads were then cooled to room temperature and collected with a magnet. The supernatant was transferred to a fresh tube and mixed with 15 μL 10 mM Tris-HCl pH 8.0.
The optimal number of PCR cycles for library amplification was determined by setting up an analytical qPCR reaction: 1 µL water, 1 µL eluted DNA, 1 µL primer mix (2.5 µM NEB Universal PCR Primer, 2.5 µM NEB Index Primer, 2.5 mM MgCl₂, 5x SYBR Green I, 5x ROX), 2.5 µL NEBNext Ultra II Q5 Master Mix. The sample was thermocycled on a qPCR machine at 98° C for 30 seconds followed by 25 cycles of 98° C for 10 seconds, 65° C for 75 seconds. The linear Rn versus cycle number was plotted to determine the cycle number corresponding to one-third of maximum fluorescence. Final library amplification was performed by setting up the following PCR in duplicate: 10.5 µL water, 2.5 µL 10 mM MgCl₂, 1 µL 25 µM NEB Universal PCR Primer, 1 µL 25 µM NEB Index Primer, 10 µL eluted DNA, 25 µL NEBNext Ultra II Q5 Master Mix. The sample was thermocycled at 98° C for 30 seconds followed by 3 or 4 cycles (as determined above) of 98° C for 10 seconds, 65° C for 75 seconds, followed by 65° C for 5 minutes, and then held at 4° C. The amplified library was purified as in “Cleanup of PCR Amplification” in the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB), adjusting the volumes as necessary. The DNA was eluted in 17 µL 10 mM Tris-HCl pH 8.0 and 15 µL was collected as the final eluate.

DNA concentration was determined with a Qubit dsDNA HS Assay, DNA integrity determined by a Bioanalyzer High Sensitivity DNA Chip, and then accurately quantified using a KAPA Library Quantification Kit. DNA was paired-end sequenced on an Illumina NextSeq 500 or HiSeq X instrument.
**Hi-C Data Processing**

Data was processed as previously described (3, 4). In brief, reads were mapped to dm3/BDGP Release 5 of the *Drosophila melanogaster* genome using BWA-MEM, reads were assigned to restriction fragments, duplicates removed, reads with a MAPQ < 30 removed, and intra-fragment reads removed. The genome was then divided into equally spaced bins and the number of contacts was counted in each pair of bins.

We noticed at kilobase and sub-kilobase resolution many empty row/columns of the matrix due to restriction fragments spanning multiple bins. Therefore, for high resolution maps, to account for the uncertainty in the location of the position of the cross-link within each restriction fragment, we randomized the read position within the restriction fragment each read mapped to and then assigned the resulting contact to its respective genomic bin. This improved the quality of the maps as “missing data” was now “recovered” without changing the position of loops or TADs. All subsequent analysis was performed using randomized intrafragment read positions. Based on the previously established metric for Hi-C map resolution (3), the resolution of our Hi-C contact maps is 260 bp. However, this approaches the median restriction fragment length for DpnII (193 bp) and, consequently, may overestimate the true resolution. For the dm3 reference genome 77.1% of restriction fragments are shorter than 500 bp, which indicates an appropriate lower bound for map resolution while at the same time recovering high-resolution information from deeply sequenced *Drosophila* Hi-C libraries.
Therefore, using the previously defined metric (3), we have attained the maximum possible resolution using DpnII for fragmentation or “sub-kilobase” resolution.

Hi-C contact maps were normalized by matrix balancing using the KR normalization algorithm (3-5). All subsequent analysis was performed on KR normalized contact maps. Reproducibility between biological replicates was determined by flattening the Hi-C matrices to vectors and calculating the Pearson's correlation coefficient ($r$) between the vectors. Biological replicates were highly correlated at all resolutions (Fig. S1D). Therefore, datasets were merged after filtering, intrafragment read positions randomized as described above, and the combined contact map KR normalized. All subsequent analysis was performed on the combined, KR normalized contact map.

TADs were identified using the previously described Arrowhead algorithm (3, 4), except with the addition of a post-processing step. This step was necessary as the Arrowhead algorithm applied at 500 bp resolution skipped some obvious larger domains, whereas the Arrowhead algorithm applied at 5 kb resolution cannot identify smaller domains. Therefore, TADs were identified at 5 kb, 2 kb, 1 kb, and 500 bp resolution using the Arrowhead algorithm, merged into a single list sorted by decreasing corner score, and conflicts, defined as the boundary of one TAD being located within another TAD, resolved by using the greater corner score of any conflicting TADs. Use of the greater corner score ensures that the most prominent, high confidence TADs are identified. Although this procedure precludes the identification of nested TADs, visual inspection of
the resulting TAD annotation revealed agreement with Hi-C contact maps and is consistent with prior annotations of non-nested TADs in *Drosophila* (6, 7).

For consistency with the loop annotation (see below), we also manually annotated TADs by visual identification of squares of enhanced contact frequency that tile the diagonal using Juicebox (8) (Fig. S7A). We identified 2,492 manually-annotated TADs. 79.2% of *Drosophila* loops did not appear at manually-annotated TAD corners, and conversely, 98.9% of manually-annotated *Drosophila* TADs did not have focal peaks at their corners. Genome-wide analysis of the distance from a loop to the nearest TAD corner for both humans and *Drosophila* was generally in agreement between the manual TAD annotation (Fig. S7B) and the Arrowhead TAD annotation (Fig. 1D).

*Drosophila* loops were manually annotated by visual identification of focal peaks of contact enrichment using Juicebox (8). The number of Hi-C contacts at peak pixels was enriched relative to four local neighborhoods (donut, bottom left, horizontal, vertical; see reference 3 for full definitions) and this enrichment was significantly greater than the enrichment at a control set of randomly shuffled loops (as described below) thereby validating our manual loop annotation (Fig. S8). For all subsequent analysis, the central 1 kb of each manually-annotated loop anchor was used.

We also used the HiCCUPS algorithm (3, 4) to annotate chromatin loops. HiCCUPS was run with options -k KR -r 1000 -f 0.001 -p 10 -i 20 -t 0.02,2.5,2.5,2.5 -d 20000. We further filtered the loop list by requiring loops to be greater than 10 kb in size (due to the
strong signal along the diagonal, it is hard to unambiguously identify very small loops), to result from collapsing more than 3 enriched pixels to a single peak pixel (to eliminate single, double, or triple pixel blowouts from being called a loop), and to be on chromosome X, 2L, 2R, 3L, 3R, or 4. This resulted in 206 loops. Using the manually annotated loop list as a gold standard (3), HiCCUPS had a 75.6% false positive rate and a 57.5% false negative rate. This elevated error rate is likely due to the very small number of loops in Drosophila and due to the fact that HiCCUPS was designed to annotate loops in mammals. In mammals, there are approximately 10,000 bona fide loops (3), so 100 falsely identified loops is a 1% false positive rate. In contrast, if there are only approximately 100 bona fide loops, as is the case in Drosophila, 100 falsely identified loops is a 50% false positive rate. This makes computational annotation of Drosophila loops very challenging. Furthermore, visual inspection of the HiCCUPS-annotated loops identified that the majority of false positives were due to the small feature size of Drosophila loops and especially TADs, which made HiCCUPS loop annotation difficult at high resolution and resulted in compartment flips being falsely identified as loops. Due to the extremely high false positive rate we only report results using the manual annotation. However, Pc and Rad21 were also significantly enriched at HiCCUPS-annotated loop anchors (Fig. S9), validating conclusions from our manual annotation. As expected for any annotation with many false positives, the enrichment of Pc and Rad21 at HiCCUPS-annotated loop anchors was less than that for manually-annotated loop anchors.

**Relationships between Loops and TADs**
To determine if loops are spatially close to TADs, we determined the Euclidean distance (i.e. the two-dimensional distance between pixels $i_1,j_1$ and $i_2,j_2$ in the Hi-C matrix; see reference 2) between a loop (i.e. focal peak of contact enrichment) and the closest TAD corner (smallest Euclidean distance) for all loops. Because TAD sizes differ between flies and humans, the Euclidean distance was normalized by the size of the closest TAD for each loop in the respective species. The mean and median TAD sizes for flies were 44.9 kb and 21.5 kb, respectively. The mean and median TAD sizes for humans were 258.3 kb and 185 kb, respectively. The mean and median TAD sizes for TADs closest to loops in flies were 173.7 kb and 148 kb, respectively. The mean and median TAD sizes for TADs closest to loops in humans were 115 kb.

To determine if loops overlapped TAD corners, that is if loops demarcate TADs, for every loop at location $M_{ij}$ we determined if there was a TAD corner within distance $0.2|i-j|$ of the loop (3). The procedure was repeated for every TAD with TAD corner at $M_{ij}$ to assess the overlap of TADs with loops.

**Enrichment of Loop Anchors at Non-histone Protein Binding Sites**

ChIP-seq reads were mapped to chromosomes X, 2L, 2R, 3L, 3R, and 4 of the dm3/BDGP Release 5 of the *Drosophila melanogaster* genome using bowtie2 with option --very-sensitive. Reads were filtered to include only properly paired reads and reads with a MAPQ $\geq 30$. PCR duplicates were removed with picard. MACS2 (9) was used to call peaks by running macs2 callpeak with parameters --SPMR --keep-dup -g dm -f BAMPE and signal tracks were computed by running macs2 bdgcmp with
parameter -m FE. Histone modification ChIP-seq data was processed as above, except the macs2 callpeak parameters --SPMR --keep-dup all --nomodel --nolambda --broad -g dm -f BAMPE were used.

The percentage of unique loop anchors at non-histone protein binding sites was calculated by dividing the number of loop anchors overlapping a ChIP-seq peak by the total number of loop anchors.

For each ChIP-seq dataset we iterated through the list of unique loop anchors counting how many unique loop anchors overlapped with a ChIP-seq peak. We then used bootstrapping to estimate the expected random distribution of counts and to get the significance of enrichment of unique loop anchors at ChIP-seq peaks. This is done by shuffling the order of unique loop anchors while ensuring that anchors remain on their respective chromosome, are not merged in the shuffle, and are excluded from genome assembly gaps. Then, we counted how many shuffled, unique loop anchors overlapped with a ChIP-seq peak. This procedure was repeated 10,000 times generating a new count each time. Enrichments and 95% confidence intervals are then determined by comparing the observed count of unique loop anchors overlapping ChIP-seq peaks to the distribution of shuffled counts.

**Enrichment/Depletion of Loops and Loop Anchors Within Epigenetic Classes**

For each epigenetic class as defined in reference 10 we iterated through the list of loops and unique loop anchors counting how many loops/anchors overlapped in their entirety
with each epigenetic class. Enrichment/depletion of loops/anchors in each epigenetic class was determined by counting how many unique loops/anchors overlapped in their entirety with an epigenetic class relative to a random shuffle control of loops/anchors using the bootstrapping approach described above, except that a Z-score and p-value for each epigenetic class was used to assess enrichment/depletion by comparing the observed count of unique loops/anchors in each epigenetic class to the distribution of shuffled counts.

The percentage of loops/anchors lying within each epigenetic was calculated by dividing the number of loops/anchors within each epigenetic class by the total number of loops/anchors. Only loops/anchors where the entire loop/anchor overlapped an epigenetic class were considered. Since many loops/anchors overlap two or more epigenetic classes the total percentage of loops/anchors lying within each epigenetic class does not sum to 100%.

**PHO and GAGA Motifs at PRC1 Loop Anchors**

PHO and GAGA motif position weight matrices were based on (11). Motifs were identified using STORM (12) and the percentage of PRC1 loop anchors containing both PHO and GAGA motifs was calculated by dividing the number of PRC1 loop anchors overlapping both a PHO and GAGA motif with the total number of PRC1 loop anchors. 10,000 random shuffles of PRC1 loop anchors, similar to that described above, was used as a control. Unique PRC1 loop anchors were identified as those loop anchors overlapping a Pc ChIP-seq peak with 30-fold or greater ChIP enrichment.
**Gene Expression at Loop Anchors**

RNA-seq data from modENCODE (13) (modENCODE accession modENCODE_4395) was downloaded from release 5.57 of FlyBase (ftp://ftp.flybase.net/releases/FB2014_03/precomputed_files/genes/gene_rpkm_report_f b_2014_03.tsv.gz). RPKM values were extracted for protein-coding genes (13,931 in total) from the Kc167 cell line dataset (FlyBase ID: FBlc0000269).

The percentage of unique PRC1 loop anchors at gene promoters (defined as the set of 1 kb windows centered on the transcription start site for each gene) was calculated by dividing the number of unique PRC1 loop anchors overlapping a gene promoter by the total number of unique PRC1 loop anchors.

Significant differences between RNA-seq RPKM values between PRC1 loop-anchor promoters, promoters bound by PRC1 not at loop anchors, promoters at loop anchors not bound by PRC1, and promoters not at loop anchors and not bound by PRC1 were compared using a one-sided Mann–Whitney U-test.

GO term enrichment was performed using DAVID (14) 6.8 Beta (May 2016 knowledgebase) with default parameters. Only GO terms with a Benjamini corrected P-value less than or equal to $10^{-3}$ were considered. REVIGO (15) was used to remove redundant GO terms with default parameters except a cut-off value (C) of 0.5 was used and the size of the GO term database was set using the *Drosophila melanogaster* GO
annotation database. Reduced redundancy GO term analysis is shown in Fig. 3D, complete GO term analysis is shown in Extended Data Table 3. No cellular component GO terms were significantly enriched.
Fig. S1. Quality assessment and reproducibility of sub-kilobase resolution *Drosophila* Hi-C.

**(A)** Distribution of four read orientations (inner, outer, left, right) as a function of distance between reads.

**(B)** Number of reads at the distance indicated on the abscissa to the closest DpnII site.

**(C)** Number of intrachromosomal contacts separated by the distance given on the abscissa.

**(D)**

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Pearson’s $r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500 bp</td>
<td>0.9060</td>
</tr>
<tr>
<td>1 kb</td>
<td>0.9591</td>
</tr>
<tr>
<td>2 kb</td>
<td>0.9827</td>
</tr>
<tr>
<td>5 kb</td>
<td>0.9948</td>
</tr>
<tr>
<td>10 kb</td>
<td>0.9974</td>
</tr>
<tr>
<td>25 kb</td>
<td>0.9996</td>
</tr>
<tr>
<td>50 kb</td>
<td>0.9998</td>
</tr>
<tr>
<td>100 kb</td>
<td>0.9999</td>
</tr>
</tbody>
</table>
(D) Pearson’s $r$ between biological replicates at the indicated Hi-C contact map resolution. Replicates were highly correlated at all resolutions.
Fig. S2. Non-histone protein ChIP-seq profiles around Drosophila loops. Mean ChIP-seq profile (upper panels) and heatmap of the ChIP-seq profile of each loop along its length (bottom panels). Loops were normalized to the same length and 25 kb of flanking DNA is shown next to each normalized loop.
Fig. S3. *Drosophila* insulators BEAF-32, Su(Hw), and CP190 are not strongly enriched at loop anchors.

(A) Hi-C contact map at 500 bp resolution reproduced from Fig. 2A of a region of chromosome X shows chromatin loops (cyan circles). CP190, Su(Hw), and BEAF-32 ChIP-seq profiles are aligned above the map. TADs are indicated by yellow outlines.

(B) Hi-C contact map at 500 bp resolution reproduced from Fig. 2B of a region of chromosome 2L shows a small network of chromatin loops (cyan circles). CP190, Su(Hw), and BEAF-32 ChIP-seq profiles are aligned above the map. TADs are indicated by yellow outlines.

(C) Left, fold enrichment of loop anchors at BEAF-32 ChIP peaks (red) and percentage of loop anchors bound by BEAF-32 (blue) at the BEAF-32 minimum ChIP enrichment indicated on the abscissa. Right, fold enrichment of BEAF-32 ChIP peaks at loop anchors (red) and percentage of BEAF-32 ChIP peaks at loop anchors (blue) at the BEAF-32 minimum ChIP enrichment indicated on the abscissa. Red shading indicates 95% confidence interval.

(D) Same as (C) except for Su(Hw).

(E) Same as (C) except for CP190.
Fig. S4. Validation of PRC1 subunit Pc at Loop Anchors.

(A) Hi-C contact map at 500 bp resolution reproduced from Fig. 2A of a region of chromosome X shows chromatin loops (cyan circles) that align with Pc ChIP-seq peaks. Pc ChIP-seq profile from an antibody different from that in Fig. 2A is aligned above the map. TADs are indicated by yellow outlines.
(B) Hi-C contact map at 500 bp resolution reproduced from Fig. 2B of a region of chromosome 2L shows a small network of chromatin loops (cyan circles) that align with Pc ChIP-seq peaks. Pc ChIP-seq profile from an antibody different from that in Fig. 2B is aligned above the map. TADs are indicated by yellow outlines.

(C) Hi-C contact map at 1 kb resolution reproduced from Fig. 3A shows chromatin loops (cyan circles) at the ANT-C Hox gene complex that align with Pc ChIP-seq peaks. Pc ChIP-seq profile from an antibody different from that in Fig. 3A is aligned above the map. TADs are indicated by yellow outlines.

(D) Hi-C contact map at 500 bp resolution reproduced from Fig. 3B shows chromatin loops (cyan circles) at the inv and en promoters that align with Pc ChIP-seq peaks. Pc ChIP-seq profile from an antibody different from that in Fig. 3B is aligned above the map. TADs are indicated by yellow outlines.

(E) Left, fold enrichment of loop anchors at Pc ChIP peaks (red) and percentage of loop anchors bound by Pc (blue) at the Pc minimum ChIP enrichment indicated on the abscissa. Right, fold enrichment of Pc ChIP peaks at loop anchors (red) and percentage of Pc ChIP peaks at loop anchors (blue) at the Pc minimum ChIP enrichment indicated on the abscissa. Pc ChIP peaks are from ChIP-seq data from an antibody different from that in Fig. 2F. Red shading indicates 95% confidence interval.
**Fig. S5. Enrichment of PRC1 loop anchors at GAGA and PHO motifs.**

(A) GAGA motif based on the position weight matrix from reference 11.

(B) PHO motif based on the position weight matrix from reference 11.

(C) Enrichment of PRC1 loop anchors at GAGA and PHO motifs. Significance of the enrichment was determined using a Fisher exact test with the mean number of 10,000 randomly shuffled PRC1 loops anchors overlapping the indicated motif as the control.
Fig. S6. Enrichment of PRC1 subunit Pc at an Independent Set of Loop Anchors.

Left, fold enrichment of loop anchors at Pc ChIP peaks (red) and percentage of loop anchors bound by Pc (blue) at the Pc minimum ChIP enrichment indicated on the abscissa. Right, fold enrichment of Pc ChIP peaks at loop anchors (red) and percentage of Pc ChIP peaks at loop anchors (blue) at the Pc minimum ChIP enrichment indicated on the abscissa. Loop anchors are from reference 16 and were converted from the dm6 to dm3 genome assembly using the FlyBase *Drosophila* Sequence Coordinates Converter. Red shading indicates 95% confidence interval.
Fig. S7. Relationship between loops and manually-annotated TADs.
(A) Hi-C contact maps reproduced from Fig. 1, 2, and 3 showing chromatin loops (cyan circles) and manually-annotated TADs (yellow outlines).

(B) Cumulative distributions of the two-dimensional Euclidean distance between a loop and the closest manually-annotated TAD corner for Drosophila loops (red), human loops (blue), and the mean of 10,000 shuffled sets of Drosophila loops (black) with 95% confidence interval (gray shading). Drosophila loops are farther from manually-annotated TAD corners than human loops ($P = 2.39 \times 10^{-7}$; two-sided KS test), but closer to manually-annotated TAD corners than the shuffled control ($P = 2.50 \times 10^{-5}$; two-sided KS test). Distance is normalized for each loop by the size of the closest TAD in the respective species because TAD sizes differ between flies and humans.
Fig. S8. Enrichment of Hi-C Contacts at Loops.

Box plots of the number of Hi-C contacts at manually annotated loops compared to the number of Hi-C contacts in four local neighborhoods (donut, bottom left, horizontal, vertical; see reference 3 for full definitions). Enrichment at manually annotated loops was significantly greater than the enrichment at a control set of randomly shuffled loops. Significance computed using a one-sided Mann–Whitney U-test.
Fig. S9. Enrichment of Pc and Rad21 at HiCCUPS-Annotated Loop Anchors.

(A) Left, fold enrichment of loop anchors at Pc ChIP peaks (red) and percentage of loop anchors bound by Pc (blue) at the Pc minimum ChIP enrichment indicated on the abscissa. Right, fold enrichment of Pc ChIP peaks at loop anchors (red) and percentage of Pc ChIP peaks at loop anchors (blue) at the Pc minimum ChIP enrichment indicated on the abscissa. Loop anchors are from loops identified by HiCCUPS (see SI Appendix, Supplementary Materials and Methods). Red shading indicates 95% confidence interval.

(B) Same as (A) except for Rad21.
Table S1. Hi-C sequencing statistics and quality metrics.

<table>
<thead>
<tr>
<th></th>
<th>Primary</th>
<th>Replicate</th>
<th>Merged</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sequencing Statistics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequenced Reads</td>
<td>532,673,670</td>
<td>554,361,974</td>
<td>1,087,035,644</td>
</tr>
<tr>
<td><strong>Alignment Statistics</strong>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal Paired</td>
<td>395,699,521 (74.29%)</td>
<td>408,451,484 (73.68%)</td>
<td>804,151,005 (73.98%)</td>
</tr>
<tr>
<td>Chimeric Paired</td>
<td>87,524,998 (16.43%)</td>
<td>92,918,735 (16.76%)</td>
<td>180,443,733 (16.60%)</td>
</tr>
<tr>
<td>Unmapped</td>
<td>10,388,903 (1.95%)</td>
<td>10,203,756 (1.84%)</td>
<td>20,592,659 (1.89%)</td>
</tr>
<tr>
<td>Ligation Motif Present</td>
<td>247,207,877 (46.41%)</td>
<td>264,925,937 (47.79%)</td>
<td>512,133,814 (47.11%)</td>
</tr>
<tr>
<td><strong>Complexity Statistics</strong>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alignable (Normal+Chimeric)</td>
<td>483,224,519 (90.72%)</td>
<td>501,370,219 (90.44%)</td>
<td>984,594,738 (90.58%)</td>
</tr>
<tr>
<td>Unique Reads</td>
<td>408,175,481 (76.63%)</td>
<td>424,129,873 (76.51%)</td>
<td>832,305,354 (76.57%)</td>
</tr>
<tr>
<td>PCR Duplicates</td>
<td>74,802,581 (14.04%)</td>
<td>77,032,979 (13.90%)</td>
<td>151,835,560 (13.97%)</td>
</tr>
<tr>
<td>Optical Duplicates</td>
<td>246,457 (0.05%)</td>
<td>207,367 (0.04%)</td>
<td>453,824 (0.04%)</td>
</tr>
<tr>
<td><strong>Unique Read Statistics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intra-fragment Reads</td>
<td>25,145,283 (4.72% / 6.16%)</td>
<td>25,523,374 (4.60% / 6.02%)</td>
<td>50,668,657 (4.66% / 6.09%)</td>
</tr>
<tr>
<td>Below MAPQ30</td>
<td>123,173,623 (23.12% / 30.18%)</td>
<td>129,753,530 (23.41% / 30.59%)</td>
<td>252,927,153 (23.27% / 30.39%)</td>
</tr>
<tr>
<td><strong>Hi-C Read Statistics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hi-C Contacts</td>
<td>259,856,575 (48.78% / 63.66%)</td>
<td>268,852,969 (48.50% / 63.39%)</td>
<td>528,709,544 (48.64% / 63.52%)</td>
</tr>
<tr>
<td>Ligation Motif Present</td>
<td>95,704,474 (17.97% / 23.45%)</td>
<td>101,798,673 (18.36% / 24.00%)</td>
<td>197,503,147 (18.17% / 23.73%)</td>
</tr>
<tr>
<td>3' Bias (Long Range)</td>
<td>69% - 31%</td>
<td>70% - 30%</td>
<td>70% - 30%</td>
</tr>
<tr>
<td>Pair Type % (L-I-O-R)</td>
<td>25% - 25% - 25% - 25%</td>
<td>25% - 25% - 25% - 25%</td>
<td>25% - 25% - 25% - 25%</td>
</tr>
<tr>
<td>Inter-chromosomal</td>
<td>12,056,707 (2.26% / 2.95%)</td>
<td>12,038,653 (2.17% / 2.84%)</td>
<td>24,095,360 (2.22% / 2.90%)</td>
</tr>
<tr>
<td>Intra-chromosomal</td>
<td>247,799,868 (46.52% / 60.71%)</td>
<td>256,814,316 (46.33% / 60.55%)</td>
<td>504,614,184 (46.42% / 60.63%)</td>
</tr>
<tr>
<td>Short Range (&lt; 20 kb)</td>
<td>107,655,986 (20.21% / 26.37%)</td>
<td>114,484,535 (20.65% / 26.99%)</td>
<td>222,140,521 (20.44% / 26.69%)</td>
</tr>
<tr>
<td>Long Range (&gt; 20 kb)</td>
<td>140,138,997 (26.31% / 34.33%)</td>
<td>142,325,315 (25.67% / 33.56%)</td>
<td>282,464,312 (25.98% / 33.94%)</td>
</tr>
</tbody>
</table>

* % Sequenced Reads
† % Sequenced Reads / % Unique Reads

For definitions of each metric see reference 3.
Table S2. External datasets used in this study.

<table>
<thead>
<tr>
<th>External Dataset</th>
<th>Reference</th>
<th>Accession Number</th>
<th>Antibody (if given/applicable)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kc167 5 chromatin classes</td>
<td>10</td>
<td>GSE22069</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kc167 BEAF-32 ChIP-Seq</td>
<td>17</td>
<td>GSE30740</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kc167 CP190 ChIP-Seq</td>
<td>17</td>
<td>GSE30740</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kc167 CTCF ChIP-Seq</td>
<td>17</td>
<td>GSE30740</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kc167 H3K27me3 ChIP-Seq</td>
<td>18</td>
<td>GSE37444</td>
<td>Millipore Cat. # 07-449 Lot # JBC1924326</td>
<td></td>
</tr>
<tr>
<td>Kc167 Pc ChIP-Seq</td>
<td>19</td>
<td>GSE63518</td>
<td>VP</td>
<td>Fig. 2, 3, S2, S5, S6, S9</td>
</tr>
<tr>
<td>Kc167 Pc ChIP-Seq</td>
<td>19</td>
<td>GSE63518</td>
<td>RJ</td>
<td>Fig. S2, S4</td>
</tr>
<tr>
<td>Kc167 Rad21 ChIP-Seq</td>
<td>18</td>
<td>GSE63518</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kc167 RNA-seq</td>
<td>13</td>
<td>modENCODE_4395</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kc167 Su(Hw) ChIP-Seq</td>
<td>17</td>
<td>GSE30740</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kc167 loops (independently annotated)</td>
<td>16</td>
<td>GSE80702</td>
<td></td>
<td>Fig. S6</td>
</tr>
<tr>
<td>GM12878 Hi-C loops</td>
<td>3</td>
<td>GSE63525</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM12878 Hi-C TADs (contact domains)</td>
<td>3</td>
<td>GSE63525</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Dataset S1. Kc167 TAD coordinates.

Juicebox format. dm3 genome assembly.

See the Excel file with the online version of this paper.

Dataset S2. Kc167 loop coordinates.

Juicebox format. dm3 genome assembly.

See the Excel file with the online version of this paper.

Dataset S3. Complete GO terms for PRC1 loop anchors.

See the Excel file with the online version of this paper.
REFERENCES


