Appendix

TADs and chromatin loops depend on cohesin and are regulated by CTCF, WAPL and PDS5 proteins

Wutz et al.

Contents:

Appendix Material and Methods

Primers used for Hi-C quality control

For quality control, candidate 3C interactions were assayed by PCR using primers listed below. The efficiency of biotin incorporation was assayed by amplifying a 3C ligation product (primers available upon request), followed by digest with *HindIII* or *NheI*.

A) short-range AHF (please reference Belton et al., 2012)
Dekker AHF64: GCATGCATTAGCCTCTGCTGTTCTCTGAAATC (chromosome 11 ; + ; 116803960-116803991)
Dekker AHF66: CTGTCCAAGTACATTCCTGTTCACAAACCC (chromosome11; + ; 116810219-116810248)

B) mid-range Myc locus
Myc locus: GGAGAACCGGTAATGGCAAA (chromosome 8; - ; 127733814-127733833)
Myc -513: GCATTCTGAAACCTGAATGCTC (chromosome 8; + ; 127220685-127220706)

C) long-range Myc locus
Myc locus: GGAGAACCGGTAATGGCAAA (same as above)
Myc +1820: AAAATGCCCATTTCCTTCTCC (chromosome 8; + ; 129554527-129554547)

Segmentation of nuclear mass and quantification of volume and intensity distribution

The nuclear mass was segmented in two sequential steps from the mRaspberry-H2B channel using a fully automated script developed in MATLAB (The MathWorks Inc., Natick, MA). In the first step, the nuclear region of interest is detected independently at each time point. In this process, a Gaussian filter of kernel size 3 with standard deviation 1.5 is applied first to each z-slice of the original image to reduce the effect of noise. Each of the original z-slices is binarized by combining two threshold values computed adaptively from the slice (2D) and from the entire stack (3D) as described in (Heriche *et al*, 2014). The nuclear mass of interest was detected from the binarized images by connecting component analysis. The nuclear mass detected in the first step contains almost the entire nuclear volume, thus it may contain low intensity regions including nucleoli. In the second step, to detect nucleus structural changes sensitively, the initial nuclear mass is segmented into high and low intensity voxels, of which the high intensity voxels are retained for further analysis. The first time point is segmented in similar fashion using both 2D and 3D threshold values determined adaptively from the histogram constructed from the voxels inside the initial nuclear mass. The total intensity of the re-segmented region in the first frame is taken as a reference to guide the segmentation of the stacks at later time points in order to obtain the same amount of total intensity inside the nuclear mass. Then, the volume of the refined segmented nuclear region is used for quantification where the volume in the first time point is normalized to 1. The higher intensity voxels in the initial segmented nuclear mass are also analyzed independently by clustering the voxels into two classes based on their intensity value. In this process, the “first class” consists of the brightest voxels that add up 50% of the total intensity inside the nucleus and the remaining voxels form the “second class”. Next, the volume of each group of voxels is computed and the ratio of the volume of the brighter first

class to that of the dimmer second class is used for quantification. The ratio obtained in the first frame is normalized to 1.

Topological domain analysis

Topologically associating domains (TADs) were identified using HOMER v4.7. We computed directionality indices (Dis) (Dixon *et al*, 2012) of Hi-C interactions in 25kb sliding windows every 5kb steps, taking into account contacts to loci 1Mb upstream and downstream from the center of the 25kb window, and smoothed the DI using a running average over a +/-25kb window. TADs were called between pairs of consecutive local maxima (start of a TAD) and minima (end of a TAD) of the smoothed DIs with a standard score difference (TAD ΔZ score) above 2.0, and the TAD ends were extended outward to the genomic bins with no directionality bias. We note that we used standardized Dis to call TADs, which is important for the reproducibility of TAD calling, especially when biological replicates might have a different amount of technical noise. On the other hand, weak DI signals due to genuine biological noise will be magnified by this approach, and might end up with an artificially large number of TADs called such as in pro-metaphase. In this case, computational detectability does not necessarily mean the presence of strong TADs. Hence, we also report the average boundary strength for each sample. To measure the average strength of TAD boundaries, we computed an average insulation score profile at the TAD boundaries. The insulation score is the standardized -log enrichment of contacts between the downstream and upstream 300kb regions (-log (*a* / (*a*+*b1*+*b2*)) where *a* is the number of contacts between, and *b1* and *b2* the number of contacts within the upstream and downstream 300kb regions). Using this definition, a more positive insulation score indicates a stronger TAD boundary. Furthermore. HOMER calls TADs as well as TAD-less regions, and when the domain structure is weaker, less of the genome would be covered by TADs. As a result, the genome coverage of the called TADs provides a further estimate of domain organization strength. We note that because of the computational detectability of weak structures, and some potential artefacts such as non-mappable regions that may appear as artificial TAD boundaries, the genome coverage, as well as the average boundary strength will not go down to zero. However, the trends are clearly seen, and the TAD boundary strength in all samples correlates well with the genome coverage of called TADs (Pearson R = 0.82) as well as with the boundary strength of TADs called in the G1 RNAi control (Pearson R = 0.88). For conditions with similar genome coverages (the WAPL, PDS5A/B, WAPL/PDS5A/B RNAi samples and their RNAi control in G1, S, G2 and pro-M cell cycle phases), we could compare the numbers and sizes of TADs. To compare contact frequencies within and around TADs, we performed aggregate TAD analysis: as an example, we plotted the average coverage and coverage-and-distance corrected Hi-C matrices around the 166 500-550kb long TADs (other TAD sizes gave the same result, data not shown). To measure changes in the average strength of TAD boundaries, for each sample we also computed an average insulation score profile in a 600kb window centered around the TAD boundaries called in the merged G1 RNAi control samples. As control, we calculated the insulation score profile around a set of control genomic positions, obtained by shifting the TAD boundaries by +1Mb.

Loop analyses

We identified loops genome-wide using the *HiCCUPS* algorithm of *Juicer tools* software (Durand *et al*, 2016). We called loops at 5kb, 10kb, and 25kb resolutions, employing Knight-Ruiz (KR) balancing and the default parameter values and a FDR threshold of 0.1, and merged these loop sets. Loops longer than 6Mb, caused by HeLa vs. hg19 assembly

mapping artifacts such as translocations, were discarded. We note that the total number of loops depends on the number of replicates used, hence we only compared experiments with the same number of replicates.

We compared the total number of loops and their size distributions in the various conditions. We identified loop anchors by any genomic loci identified at a loop end with overlapping genomic coordinates, and calculated the frequency distribution of the number of loops from these loop anchors. Using the loop anchors and all loops connecting them, we defined connected networks (components, in graph theory) among all loops that are not connected by a loop to any outside genomic locus, and calculated the frequency distribution of loop numbers in these isolated networks of loops.

For measuring the aggregate peak enrichment of our loop sets, we used aggregate peak analysis. For loops in the 750kb-6Mb range, we used the *APA* algorithm of *Juicer tools* v0.7.5 to plot the sum of coverage-corrected Hi-C sub-matrices at 25kb resolution (Rao *et al*, 2014). The sub-matrices were centered and aligned around the peak coordinates of the looping loci such that the upstream loop anchor was at the center of the vertical, and the downstream loop anchor was at the center of the horizontal axis. The resulting plot displays a measure for the number of contacts that lie within the entire putative peak set at the center of the matrix, and the aggregation of contacts in a focal enrichment compared to the surroundings. The bottom left quarter of the plot displays the contact counts between loci between the looping anchors, characteristic of TAD-forming loops, and there is a general contact count decrease from the bottom left corner to the top right corner reflecting the distance dependent contact decay characteristic of the Brownian motion of linear polymers. To decouple this distance dependency, for loops that were exactly 300kb or 600kb long, we also plotted the average coverage-and-distance corrected Hi-C sub-matrices at 5kb resolution around the loops, using *HOMER*. Matrices were aligned as before, such that the diagonal of the sub-matrix also overlaps with the main diagonal of the Hi-C matrix. The plots show the contact enrichment compared to random contacts due to Brownian motion of the polymer. The selection of specific loop lengths enables more striking visualization of the domains bordered by some of the loops in the set. Another advantage of using the average rather than the sum of contacts is that it corrects for count differences due to different numbers of loops in the datasets, making the plots comparable between different loop sets.

To address the violation of the CTCF convergence rule in our datasets, we identified loops for which both anchors had consensus CTCF binding motifs overlapping with CTCF ChIP-seq peaks, as well as at least one SMC3 ChIP-seq peak. For these loops, we calculated the frequency of convergent, tandem and divergent CTCF sites in the control dataset, as well as for loops identified in WAPL, PDS5A/PDS5B and joint depleted datasets but not in control. For chains of loops, in which downstream loop anchors are also upstream loop anchors, we divided the loops into 5', internal and 3' loop categories (a 5' loop's upstream loop anchor only interacts downstream, and a 3' loop's downstream loop anchor only interacts upstream), and counted the frequency of forward and reverse CTCF binding motifs at these loop anchors.

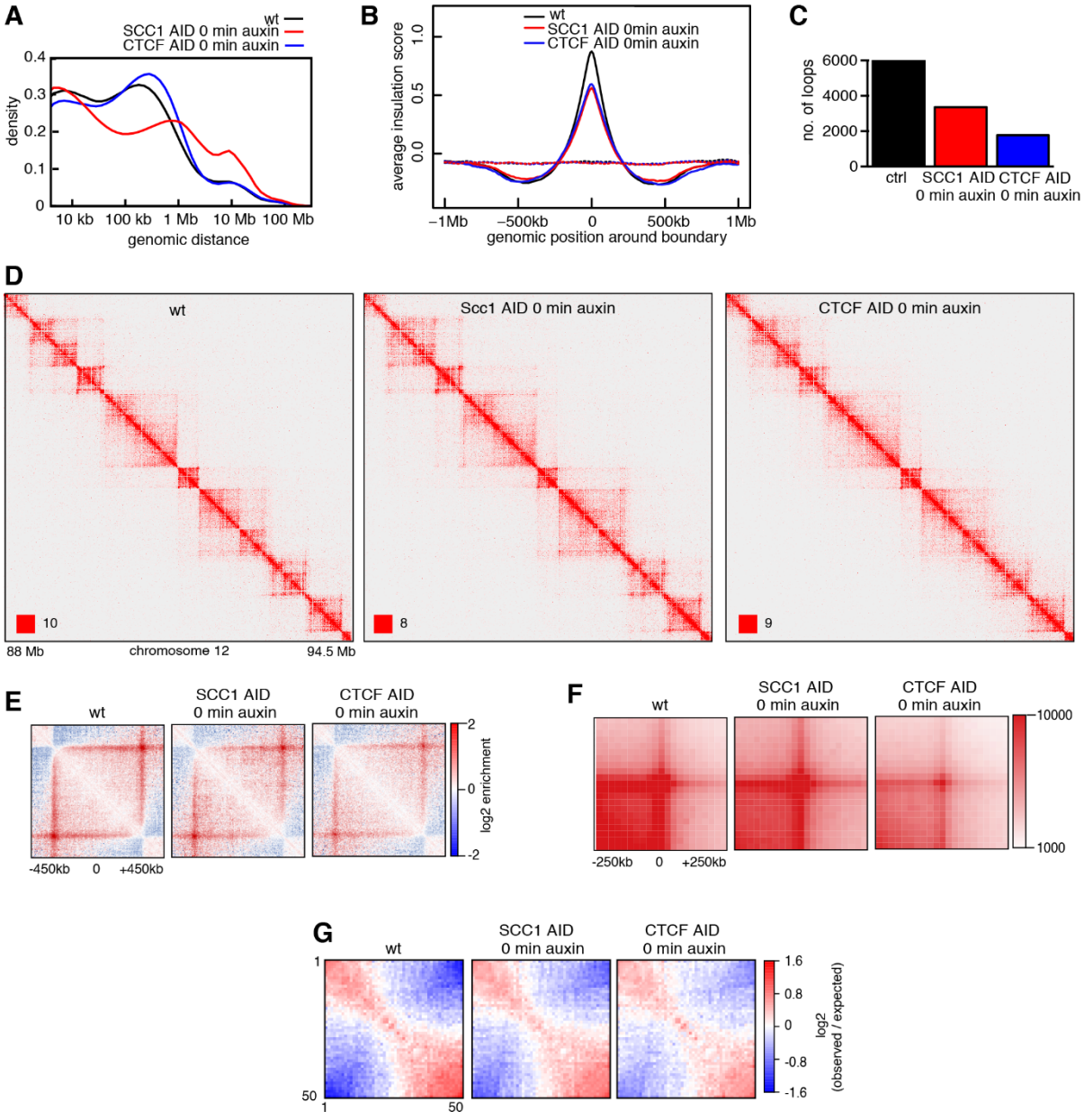
Appendix References

Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y & Dekker J (2012) Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* **58:** 268–276

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES & Aiden EL (2016) Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst.* **3:** 95–98

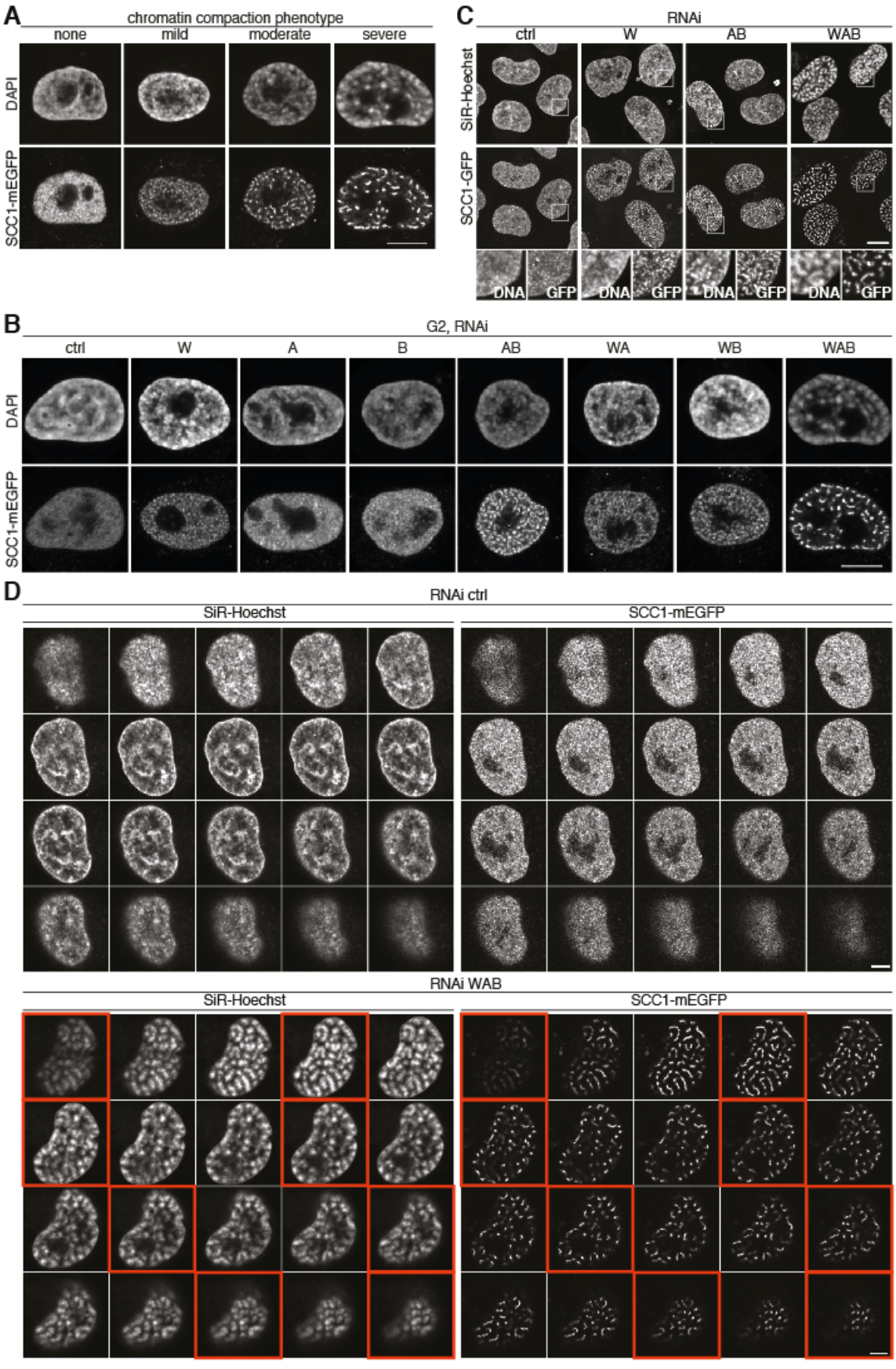
Heriche J-K, Lees JG, Morilla I, Walter T, Petrova B, Roberti MJ, Hossain MJ, Adler P, Fernandez JM, Krallinger M, Haering CH, Vilo J, Valencia A, Ranea JA, Orengo C & Ellenberg J (2014) Integration of biological data by kernels on graph nodes allows prediction of new genes involved in mitotic chromosome condensation. *Mol. Biol. Cell* **25:** 2522–2536

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES & Aiden EL (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159:** 1665–1680



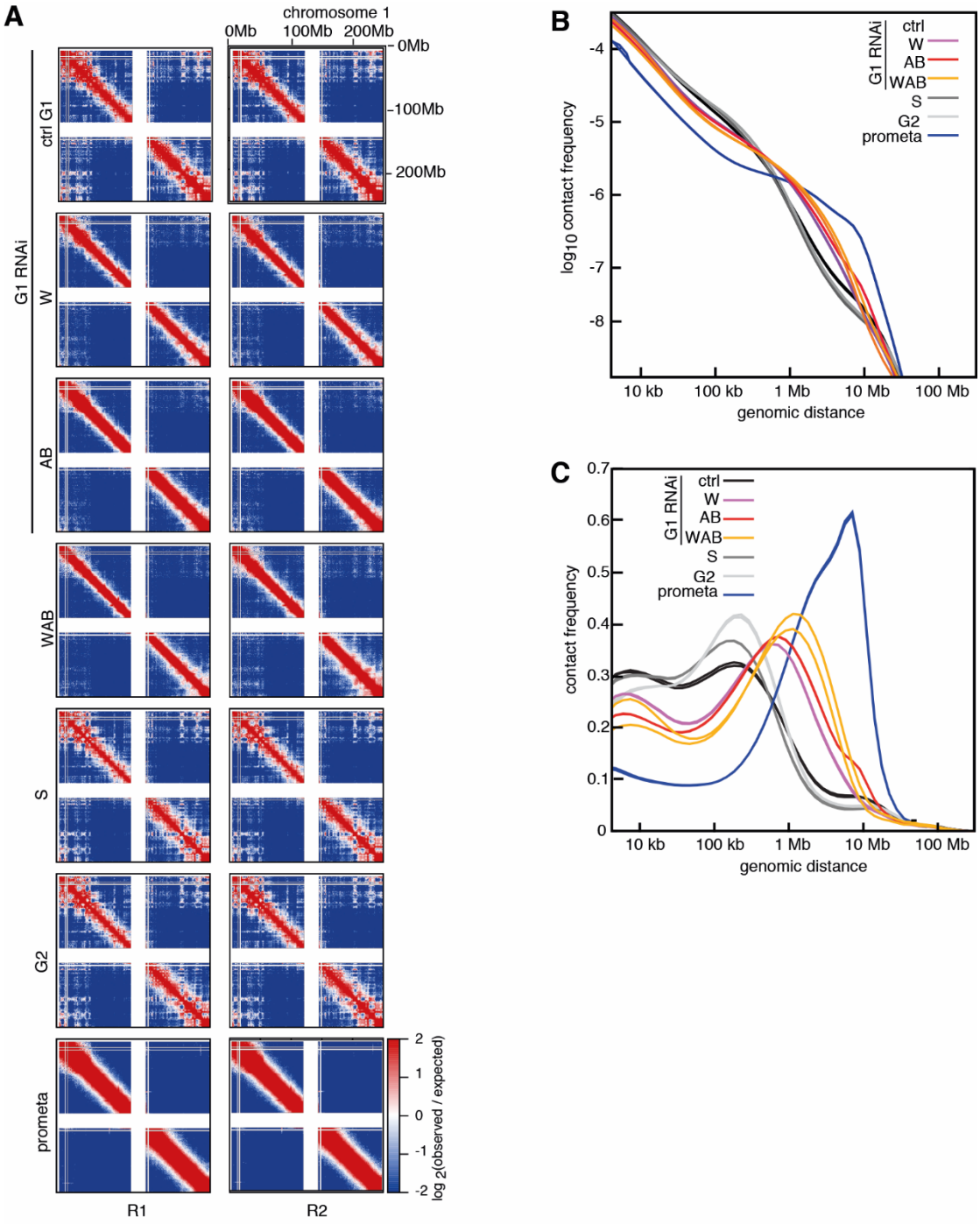
Appendix Figure S1 related to Figures 1 and 2.

(A) Intra-chromosomal contact frequency distribution as a function of genomic distance, in the wild type (WT, black), SCC1-mEGFP-AID (red) and CTCF-mEGFP-AID (blue) cells immediately after auxin addition. **(B)** Average insulation score around TAD boundaries identified in G1 control cells, for the same conditions as in (A). Dashed lines show the average insulation score around the +1Mb-shifted boundaries as control. **(C)** Number of loops identified by HiCCUPS, for the same conditions. **(D)** Coverage-corrected Hi-C contact count matrices in the 88-94.5Mb region of chromosome 12, for the same conditions. **(E)** Average contact enrichment around loops after auxin addition, for the 82 x 600 kb long loops identified by HiCCUPS in G1 control. **(F)** Total contact counts around loops after auxin addition, for all 750kb-6Mb long loops identified by HiCCUPS in G1 control. **(G)** Inter-contact enrichment between bins with varying compartment strength from most B-like (1) to most A-like (50).


A
chromatin compaction phenotype
none
mild
moderate
severe
DAPI
SCC1-mEGFP
C
RNAi
ctrl
W
AB
WAB
SiR-Hoechst
SCC1-GFP
DNA
GFP
B
G2, RNAi
ctrl
W
A
B
AB
WA
WB
WAB
DAPI
SCC1-mEGFP
D
RNAi ctrl
SiR-Hoechst
SCC1-mEGFP
RNAi WAB
SiR-Hoechst
SCC1-mEGFP

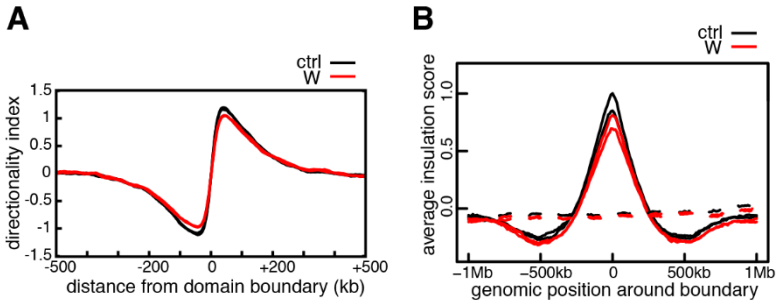
Appendix Figure S2 related to Figure 3

(A) Representative immunofluorescence images of SCC1-mEGFP cells used for phenotypic classification analysis shown in Fig 3C. Cells were stained for GFP, DNA was counterstained with DAPI. Scale bar indicates 10 μm. **(B)** Representative immunofluorescence images of SCC1-mEGFP cells in G2-phase stained for GFP. DNA was counterstained with DAPI. Scale bar indicates 10 µm.**(C)** Live cell imaging of SCC1-mEGFP cells that had been depleted for control, W, AB and WAB. DNA was stained with SiR-DNA by Spirochrome. Three of the confocal sections are shown by perspective view. Scale bar indicates 10 µm. **(D)** Live cell imaging of SCC1-mEGFP cells that were depleted for WAB or control depleted. Individual confocal sections are shown (confocal distance between original sections = 0.4 µm). The images entangled by the orange lines are also shown in Fig 3E. DNA was counter stained with SiR-DNA by Spirochrome. Scale bar indicates 5 µm.



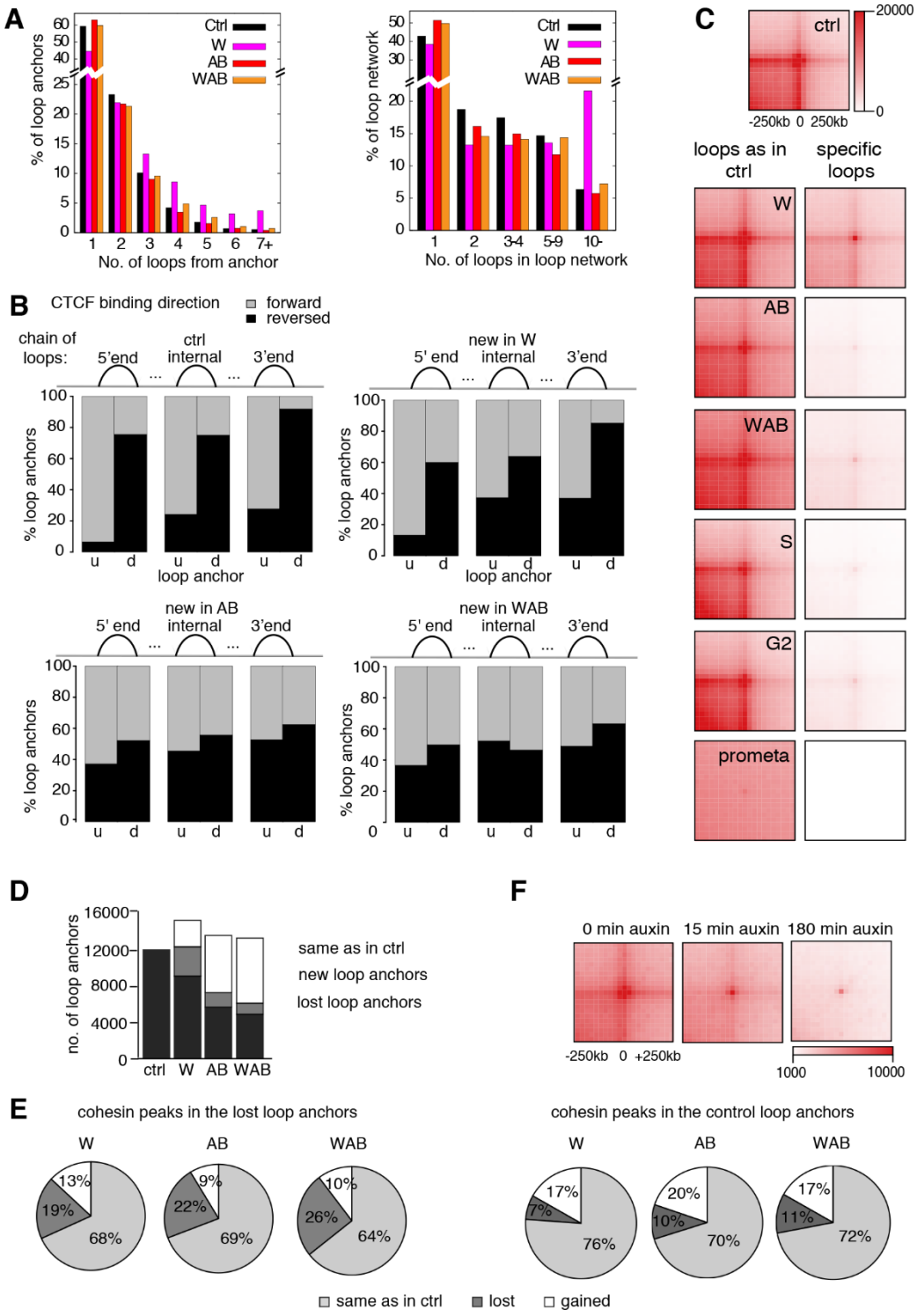
Appendix Figure S3 related to Figure 5

(A) Coverage-corrected Hi-C contact enrichment matrices (using HOMER) of chromosome 1, for the same conditions as in Fig 5, showing the two biological replicates side by side. **(B-C)** Intra-chromosomal contact frequency distribution as a function of genomic distance using equal sized (B) or logarithmically increasing (C) genomic distance bins, for the two biological replicates. Colors are the same as in Fig 5.



Appendix Figure S4

(A) Average standardized directionality index profiles in a 1Mb region centered around TAD boundaries identified in the wild-type (ctrl, black) and WAPL KO (W, red) HAP1 cells of (Haarhuis *et al*, 2017). **(B)** Average insulation score around TAD boundaries identified in the wild-type HAP1 cells, for the same conditions as in (A). Dashed lines show the average insulation score around the +1Mb-shifted boundaries as control. Colors are as in (A).


A
% of loop anchors
Ctrl
W
AB
WAB
No. of loops from anchor
% of loop network
No. of loops in loop network
B
CTCF binding direction
forward
reversed
chain of loops:
5'end
ctrl internal
3'end
new in W
new in AB
new in WAB
% loop anchors
loop anchor
C
ctrl
20000
0
-250kb 0 250kb
loops as in ctrl
specific loops
W
AB
WAB
S
G2
prometa
D
no. of loop anchors
ctrl W AB WAB
same as in ctrl
new loop anchors
lost loop anchors
F
0 min auxin
15 min auxin
180 min auxin
-250kb 0 +250kb
1000
10000
E
cohesin peaks in the lost loop anchors
cohesin peaks in the control loop anchors
W
AB
WAB
13%
19%
68%
9%
22%
69%
10%
26%
64%
17%
7%
76%
20%
10%
70%
17%
11%
72%
same as in ctrl
lost
gained

Appendix Figure S5 related to Figures 9 and 10

(A) The frequency distributions of the number of loops formed from all loop anchors (left), and the number of loops forming all contiguous loop networks (right), for the control-depleted (ctrl, black), WAPL depleted (W, magenta), PDS5A/B depleted (AB, red) and WAPL/PDS5A/B depleted (WAB, orange) RNAi samples. **(B)** For 5', internal and 3' loops of loop networks, the proportion of their upstream (u) and downstream (d) loop anchors with forward (grey) and reverse (black) CTCF binding orientation. **(C)** Total contact counts around loops for all 750kb-6Mb long loops identified by HiCCUPS in control-depleted G1 cells (left), or in the corresponding sample but not in G1 control cells (right). **(D)** Changes in loop anchors compared to control-depleted cells, showing the number of retained (black), newly appearing (grey) and lost (white) loop anchors, in the conditions listed in (A). **(E)** Changes in cohesin binding at the disappearing and retaining loop anchors shown in (D), showing the fractions of loop anchors with unchanging (light grey), lost (dark grey) and newly gained (white) cohesin peaks. **(F)** Total contact counts around loops after auxin addition in the WAPL/PDS5A/B RNAi depleted SCC1-mEGFP-AID cell line, 0, 15 and 180 minutes after auxin addition, for all 750kb-6Mb long loops identified by HiCCUPS in control-depleted G1 cells.