SUPPLEMENTARY INFORMATION

SUPPLEMENTARY METHODS

CRISPR sgRNA lentivirus production

TKOv3 library lentivirus was produced by co-transfection of lentiviral vectors psPAX2 (packaging vector, Addgene #12260) and pMD2.G (envelope vector, Addgene #12259) with TKOv3 lentiCRISPRv2 (pLCV2) plasmid library, using X-treme Gene9 transfection reagent (Roche). Briefly, HEK293T cells were seeded at a density of $9x10^6$ cells per 15cm plate and incubated overnight, after which cells were transfected with a mixture of psPAX2 (4.8µg), pMDG.2 (3.2µg), TKOv3 plasmid library (8µg), and X-treme Gene9 (48µL) in Opti-MEM (GIBCO), in accordance with the manufacturer's protocol. 24 hours after transfection, the medium was changed to serum-free, high BSA growth medium (DMEM with1% BSA (Sigma) and 1% penicillin/streptomycin (GIBCO). Virus-containing medium was harvested 48 hours after transfection, centrifuged at 500g for 5 minutes, and stored at -80°C. Functional titers in cells to be screened were determined by virus titration: 24 hours after infection, the medium was replaced with puromycin-containing medium (1mg/ml), and cells were incubated for 48 hours. The multiplicity of infection (MOI) of the titrated virus was determined 72 hours after infection by comparing the survival of infected cells to infected unselected and non-infected selected control cells.

Pooled genome-wide CRISPR screens in HAP1 CTCF degron cell lines

CRISPR TKOv3 screens were performed, as previously described[1–3], in the HAP1-CTCFdegron and HAP1-CTCFdegron-TIR1 cells. For HAP1-CTCFdegron, 54M cells were infected and for HAP1-CTCFdegron-TIR1, 110M cells were infected with the TKOv3 lentiviral library at MOI of ~0.3 for reaching >200-fold coverage of the library after puromycin selection. 24 hours after infection, the infected cells were selected by changing the medium to puromycin-containing medium (1µg/mL). 72 hours after infection was considered as T0. 30M cells were harvested and 135M puromycin-selected cells were seeded in medium without puromycin to keep the >200-fold coverage of the library for each of the three screen replicates and each of the three different conditions (non-treated (NT), 25µM auxin, 500µM auxin). At T3, 20M cells were harvested. The remaining cells were divided into three conditions (non-treated (NT), 25µM auxin, 500µM auxin). Each condition was screened in triplicates with the cells seeded at the appropriate number to keep the 200-fold coverage of the library (i.e., 15M cells per triplicate per condition). Cells were passaged every 3 days and maintained in absence (NT) or in presence of auxin (25µM, 500µM IAA) at 200-fold coverage until reaching T15. For T6, T9, T12 and T15, 20M cells were harvested for each triplicate in each condition.

Genomic DNA was extracted for T0, T6 and T15 of each triplicate of each condition using the Wizard Genomic DNA Purification kit (Promega, A1125) following manufacturer's instructions. Sequencing libraries were prepared by amplifying sgRNA inserts via a 2-step PCR. The first PCR enriches for the sgRNA regions in the genome and the second PCR adds indices by using primers that include Illumina TruSeq adaptors. Briefly, in PCR1 50µg of genomic DNA were amplified in 15 parallel reactions to maintain the 200-fold coverage. DNA was amplified using the NEBNext Ultra II Q5 Master Mix. PCR1 primers were FW1: GAGGGCCTATTTCCCATGATTC, RS1: GTTGCGAAAAAGAACGTTCACGG. For PCR2, 5µL of the pooled PCR1 product was used and amplified using the NEBNext Ultra II Q5 Master Mix and primers with i5 and i7 indices. PCR2 primers were FW2: AATGATACGGCGACCACCGAGATCTACACNNNNNNNNACACTCTTTCCCTACACGACG CTCTTCCGATCTTTGTGGAAAGGACGAAACACCG, RS2: CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCT TCCGATCTACTTGCTATTTCTAGCTCTAAAAC, with NNNNNNNN being the indices of the TruSeq adaptors. After PCR2, the 200bp amplified band was excised from an agarose gel and purified using QIAGEN MinElute Gel Extraction kit (Qiagen, 28604). Libraries were quantified using a Qubit fluorometer and pooled for sequencing. Libraries were sequenced on an Illumina HiSeq2500. Each read was completed with standard primers for dual indexing. The first 21 cycles of sequencing were dark cycles, or base additions without imaging. The actual 26-bp read begins after the dark cycles and contains two index reads, reading the i7 first, followed by i5 sequences. For the T0 libraries we aimed for a total of 30M reads each and for the T6 and T15 libraries we aimed for a total of 20M reads each. The actual number of reads can be found in the Supplementary Table 2.

ChIP-seq

Cell fixation

Cells were seeded in 15cm plates in order to have ~15M cells per antibody. Cells were fixed in the plates with 1% formaldehyde in HBSS for 10 minutes at room temperature. To quench the formaldehyde and terminate the cross-linking reaction 125mM glycine was added and incubated for 5 minutes at room temperature. Plates were then transferred and incubated on ice for 15 minutes. Formaldehyde solution was then removed, and cells were washed with dPBS. Cells were scraped from the plates in dPBS containing protease inhibitors, spun and supernatant was removed.

Cell lysis

Each 5M cells were resuspended in 1mL of hypotonic lysis buffer (20mM TRIS HCl pH8.0, 85mM KCl, 0.5% IGEPAL, 1X HALT protease inhibitor) and incubated 15 minutes on ice. Cells were then spun down 5 minutes, 500g, 4°C, supernatant was removed and 5M cells were

resuspended in 300µL of sonication buffer (20mM TRIS HCl pH8.0, 0.2% SDS, 0.5% Sodium Deoxycholate, 1X HALT protease inhibitor).

Sonication

Cells were sonicated using a bioruptor pico, at 4°C, following the manufacturer instructions and with the following settings: 30s on, 30s off, for 15 cycles. After sonication, tubes were briefly spun, combined together and spun for 10 minutes, 12,000g, 4°C. The chromatin fraction is in the supernatant.

Assessment of the sonication

To check if sonication was efficient, 10µL of each sample was combined with 200µL of de-crosslinking solution (50mM TRIS HCl, pH 8.0, 200mM NaCl, 0.2mg/mL of proteinase K) incubated overnight at 65°C. The next day de-crosslinked samples were treated with 0.1mg/mL RNAseA for 15 minutes at 37°C and DNA was extracted with phenol/chloroform in phase lock tubes (first wash: phenol chloroform, second wash: chloroform). DNA was concentrated using an amicon column and eluted in 10µL of water before loading the totality on a 1.5% agarose gel. Most of the DNA fragments are between 150-400bp.

Chromatin dilution and antibody incubation

300µL of chromatin from 5M cells was diluted with 1,200µL of dilution IP buffer (20mM TRIS HCl pH 8.0, 150mM NaCl, 2mMEDTA, 1% TRITON, 1X HALT protease inhibitor). For input control, 10% of the diluted chromatin was set aside. Antibody was added to each tube (CTCF: 20µL, RAD21: 4µL, DDX55: 4µL, TAF5L: 12µL) and incubated overnight at 4°C on a rotator.

Immunoprecipitation

20µL per tube of proteinG Dynabeads were washed in IP buffer 1X (20mM TRIS pH8, 2mM EDTA, 150mM NaCl, 1% TRITON, 0.1% SDS, 1X HALT protease inhibitor), added to the tube and incubated 2 hours at 4°C on a rotator.

To get rid of the non-specific binding, first two 5 minutes washes with 600µL of IP buffer 1X were done (20mM TRIS pH8, 2mM EDTA, 150mM NaCl, 1% TRITON, 0.1% SDS, 1X HALT protease inhibitor), second two 5 minutes washes with 600µL of wash buffer B (20mM Tris HCl, pH 8.0, 500mM NaCl, 2mM EDTA, 0.5% Na Deoxycholate, 1% Triton X-100, 1X HALT protease inhibitor), third, two 5 minutes washes with 600µL of wash buffer C (20mM Tris HCl, pH 8.0, 1mM EDTA, 0.5% Na Deoxycholate, 250mM LiCl, 1% Triton X-100, 1X HALT protease inhibitor) and finally two 5 minutes washed with 600µL of TLE (10mM Tris HCl, pH 8.0, 0.1mM EDTA, 1X HALT protease inhibitor)

Elution and DNA purification

DNA was eluted twice by adding 50µL of elution buffer (50mM NaHCO3, 1% SDS) to the beads and incubating 30 minutes at 65°C.

The saved input DNA was diluted in the same elution buffer and treated similarly.

RNA was degraded by incubating the samples with 0.1mg/mL for 30 minutes at 37°C.

Samples were de-crosslinked by adding 200mM NaCl and 0.2mg/mL of proteinase K overnight at 65°C. The next day, DNA samples were extracted with phenol/chloroform in phase lock tubes (first wash: phenol chloroform, second wash: chloroform). DNA was precipitated with 0.3M Sodium Acetate (pH 5.2), 0.1mg/ml glycogen and 2X volume of ethanol. Samples were incubated on ice for 3 hours to precipitate DNA. DNA was precipitated by spinning for 30 minutes at 16,000g at 4°C, washed once with ethanol 70%, spun 10 minutes at 16,000g, supernatant was removed, and DNA pellets were air dried and resuspended in 40µL of water. DNA concentrations were assessed by Qubit and the quality of the ChIP-seq library by qPCR.

Library preparation

End repair

For end repair, 35µL of sample were transferred to a PCR tube, then 15µL of the end-repair mix (1X NEB ligation buffer (NEB B0202S), 17.5mM dNTP mix, 7.5U T4 DNA polymerase (NEBM0203L), 25U T4 polynucleotide kinase (NEB M0201S), 2.5U Klenow polymerase Polymerase I (NEB M0210L)) were added. The reactions were then incubated at 37°C for 30 minutes, followed by incubation at 75°C for 20 minutes to inactivate Klenow polymerase. An ampure beads purification was then performed (1.6X) and DNA was eluted in 41µL of water.

A-tailing

dATP was added to the 3' ends by adding 9µL of A-tailing mix (5µL NEB buffer 2.1, 1µL of 1mM dATP, 5U Klenow exo (NEB M0212S)) to the 41µL of DNA sample from the previous step. The reaction was incubated in a PCR machine at 37°C for 30 minutes, then at 65°C for 20 minutes. An ampure beads purification was then performed (1.6X) and DNA was eluted in 40µL of water.

Illumina adapter ligation and PCR

The TruSeq DNA LT kit Set A (Illumina, #15041757) was used. 12.5µL of ligation mix (5µL Illumina paired-end adapters (non-diluted if library concentration is more than 100ng, diluted by 15 if library concentration is more than 10ng, diluted by 25 if library concentration is around 5ng), 4µL T4 DNA ligase Invitrogen, 2.5µL 5x T4 DNA ligase buffer (Invitrogen 5X), 1µL of 10mM ATP) was added to the 40µL sample from the previous step. The ligation samples were then incubated at room temperature for 2 hours. An ampure beads purification was then performed (1X) and DNA was eluted in 20µL of water. One PCR reaction (14 cycles) was then performed as follows (20µL DNA, 5µL of Primers mix (TruSeq DNA LT kit Set A 15041757), 20µL Master Mix (TruSeq DNA LT kit Set A 15041757), 5µL of water). An ampure beads purification was then performed (0.64X-1.1X) to select DNA fragments between 200-500bp to remove primers and longer amplified DNA fragments that could inhibit the sequencing reaction and DNA was eluted in 25µL of water. The libraries were sequenced using 50bp paired end reads on an Illumina HiSeq4000. Each ChIP-seq was performed in two replicates.

RNA-seq

RNA was extracted using the RNeasy Mini Kit (Qiagen, 74104) with QIAshredder (Qiagen, 79654). RNA was sent to the BGI Hong Kong facility (DNBseq Eukaryotic Transcriptome resequencing) for library preparation and sequencing on a DNBseq platform MGISEQ-G400 (100bp paired ends). Each RNA-seq library was performed in two replicates.

Hi-C

Hi-C was performed as described previously[4] with minor modifications.

Cell fixation

Cells were fixed in the plates with 1% formaldehyde in HBSS for 10 minutes at room temperature. 125mM glycine was added to the plate and cells were incubated for 5 minutes at room temperature to quench the formaldehyde and terminate the cross-linking reaction. Plates were then incubated on ice for 15 minutes. Formaldehyde solution was then removed, and cells were washed with dPBS. Cells were scraped from the plates, spun and supernatant was removed. If Hi-C was not performed immediately after cell fixation, pellets were flash frozen in liquid nitrogen and kept at -80°C until starting the cell lysis.

Cell lysis

5M formaldehyde cross-linked cells were incubated in 1,000µL of cold lysis buffer (10mM Tris-HCl pH8.0, 10mM NaCl, 0.2% (v/v) Igepal CA630, mixed with 10µL of 10X protease inhibitors (Thermofisher 78438)) on ice for 15 minutes. Next, cells were lysed with a Dounce homogenizer and pestle A (Kimble Kontes # 885303-0002) by moving the pestle up and down 30 times, incubating on ice for one minute followed by 30 more strokes with the pestle. The suspensions were centrifuged for 5 minutes at 2,000g at RT using a tabletop centrifuge (Centrifuge 5810R, (Eppendorf)). The supernatants were discarded, and the pellets were washed twice with ice cold 500µL 1x NEBuffer 3.1 (NEB). After the second wash, the pellets were resuspended in 720µL of 1x NEBuffer 3.1 and split into two tubes. 18µL were kept at -20°C to assess the chromatin integrity later. Chromatin was solubilized by addition of 38µL 1% SDS per tube and the mixture was resuspended and incubated at 65°C for 10 minutes. Tubes were put on ice and 43µL 10% Triton X-100 was added.

Chromatin digestion

Chromatin was digested by adding 400 Units DpnII (NEB) per tube at 37°C for overnight digestion with alternating rocking. Digested chromatin samples were incubated at 65°C for 20 minutes to inactivate the DpnII enzymes, spun shortly and transferred to ice. 10µL were kept at -20°C to assess the digestion efficiency later.

Biotin fill-in

DNA ends were marked with biotin-14-dATP by adding 60µL of biotin fill-in master mix (1X NEB 3.1, 0.25mM dCTP, 0.25mM dGTP, 0.25mM dTTP, 0.25mM biotin-dATP

(ThermoFisher#19524016), 50U Klenow polymerase Polymerase I (NEB M0210L)). The samples were incubated at 23°C for 4 hours with agitation and then placed on ice.

Blunt end ligation

Ligations were performed by adding 665µL of ligation mix (240µL of 5X ligation buffer (1.8X) (Invitrogen), 120µL 10% Triton X-100, 12µL of 10mg/mL BSA, 50µL T4 DNA ligase (Invitrogen 15224090), and 243µL ultrapure distilled water (Invitrogen)). The reactions were then incubated at 16°C for 4 hours with some agitation. After the ligation, the crosslink was reversed by adding 50µL of 10mg/mL proteinase K (Fisher BP1750I-400) and incubated at 65°C for 2 hours followed by a second addition of 50µL of 10 mg/mL Proteinase K and overnight incubation at 65°C.

DNA purification

Reactions were cooled to room temperature and the DNA was extracted by adding an equal volume of saturated phenol pH 8.0: chloroform (1:1) (Fisher BP1750I-400), vortexing for 1 minute, transferred to a phase-lock tube and spun at 16,000g for 5 minutes. DNA was precipitated by adding 1/10th of 3 M sodium acetate pH 5.2, 2 volumes of ice-cold 100% ethanol and incubated for at least an hour at -80°C. Next, the DNA was pelleted at 16,000g at 4°C for 30 minutes. The supernatants were discarded, the pellets were dissolved in 500µL 1X TLE and transferred to a 15mL AMICON Ultra Centrifuge filter (UFC903024 EMD Millipore). 10mL of TLE was added to wash the sample, the columns were spun at 4,000g for 10 minutes and the flowthroughs were discarded. A second wash with 10mL of TLE was done and the sample was transferred to a 0.5mL AMICON Ultra Centrifuge filter (UFC5030BK EMD Millipore) and spun at 16,000g for 10 minutes to reduce the sample to 50µL. RNA was degraded by adding 1µL of 10 mg/mL RNAase A and incubated at 37°C for 30 minutes. DNA was quantified by loading on a 1% gel 1µL of the Hi-C sample, the chromatin integrity and the digestion controls.

Biotin removal from unligated ends

To remove biotinylated nucleotides at DNA ends that did not ligate, the Hi-C samples were treated with T4 DNA polymerase. 5µg of Hi-C library were incubated with 5µL 10x NEBuffer 3.1, 0.025mM dATP, 0.025mM dGTP and 15U T4 DNA polymerase (NEB # M0203L) in 50µL. Reactions were incubated at 20°C for 4 hours, the enzymes were then inactivated at 75°C for 20 minutes and placed at 4°C.

DNA shearing

The samples were pooled, and the volume was brought up to 130µL 1X TLE. The DNA was sheared to a size of 100-300 bp using a Covaris instrument (Duty Factor 20%, Cycles per Burst 200, peak power 50, average power 17.5 and process time 180 sec). The volume was brought up to 500µL with TLE for Ampure fractionation. To enrich for DNA fragments of 100-300bp an Ampure XP fractionation was performed (Beckman Coulter, A63881) and the DNA

was eluted with 50µL of water. The size range of the DNA fragments after fractionation was checked by running an aliquot on a 2% agarose gel.

End repair

To proceed for end repair, 45µL of Hi-C sample was transferred to a PCR tube, then 25µL of the end-repair mix (3.5X NEB ligation buffer (NEB B0202S), 17.5mM dNTP mix, 7.5U T4 DNA polymerase (NEBM0203L), 25U T4 polynucleotide kinase (NEB M0201S), 2.5U Klenow polymerase Polymerase I (NEB M0210L)) was added. The reactions were then incubated at 37°C for 30 minutes, followed by incubation at 75°C for 20 minutes to inactivate Klenow polymerase.

Biotin pull down

To pull down biotinylated DNA fragments, 50µL of MyOne streptavidin C1 beads mix (Thermo Fisher 65001) was transferred to a 1.5mL tube. The beads were washed twice by adding 400µL of TWB (5mM Tris-HCl pH8.0, 0.5mM EDTA, 1M NaCl, 0.05% Tween20) followed by incubation for 3 minutes at RT. After the washes, the beads were resuspended in 400µL of 2X Binding Buffer (BB) (10mM Tris-HCl pH8, 1mM EDTA, 2M NaCl) and mixed with the 400µL DNA from the previous step in a new 1.5mL tube. The mixtures were incubated for 15 minutes at RT with rotation. The DNA bound to the beads was washed first with 400µL of 1X BB and then with 100µL of NEB2.1 1X. Finally, beads with bound DNA were resuspended in 41µL of NEB2.1 1X.

A-tailing

Then, dATP was added to the 3' ends by adding 9µL of A-tailing mix (5µL NEB buffer 2.1, 5µL of 1mM dATP, 3U Klenow exo (NEB M0212S)) to the 41µL of beads with bound DNA from the previous step. The reaction was incubated in a PCR machine (at 37°C for 30 minutes, then at 65°C for 20 minutes, followed by cooling down to 4°C). Next, the tube was placed on ice immediately. The beads with bound DNA were washed twice using 100µL 1X T4 DNA Ligase Buffer (Invitrogen). Finally, beads with bound DNA were resuspended in 36.25µL 1X T4 DNA Ligase buffer (Invitrogen).

Illumina adapter ligation and PCR

The TruSeq DNA LT kit Set A (Illumina, #15041757) was used. 10µL of ligation mix (3µL Illumina paired-end adapters, 4µL T4 DNA ligase Invitrogen, 2.75µL 5X T4 DNA ligase buffer (Invitrogen 5X)) was added to the 36.25µL beads with bound DNA from the previous step. The ligation samples were then incubated at room temperature for 2 hours on a rotator. The beads with bound DNA were washed twice with 400µL of TWB, then twice using 100µL NEB2.1 1X. Finally, the samples were resuspended in 20µL of NEB2.1 1X. Two trial PCR reactions (6 and 8 cycles) were performed as follows (0.9µL DNA bound to beads, 1.5µL of Primers mix (TruSeq DNA LT kit Set A 15041757), 6µL Master Mix (TruSeq DNA LT kit Set A 15041757), 6.6µL of ultrapure distilled water (Invitrogen)). The number of PCR cycles to generate the final

Hi-C material for deep sequencing was chosen based on the minimum number of PCR cycles in the PCR titration that was needed to obtain sufficient amounts of DNA for sequencing. ClaI digestion was done as a library quality check. A downward shift of the amplified DNA to smaller sizes indicates that DNA ends were correctly filled in and ligated (creating a ClaI site). Primers were removed using Ampure XP beads. The libraries were sequenced using 50bp paired end reads on an Illumina HiSeq4000. Each Hi-C was performed in two replicates.

CRISPR screen analysis

Library processing was performed as in[5]. Reads were trimmed by removing up to 20bp after the first 8bp of the anchors used in the barcoding primers. Reads were aligned using BOWTIE version 0.12.8 (allowing for max. 2 mismatches, ignoring qualities)[6]. Read counts for each screen library were normalized to 10M reads total per sample (see Supplementary Table 2 containing the CRISPR screen statistics). Fold change was calculated to the T0 reference sample. For precision and recall plots, we generated Bayes Factor (BF) scores with the calculated fold changes using BAGEL version 0.91[7]. We then calculated precision and recall using these BFs, discarding scores for genes that were not included in version 2 training sets (union of essential (EG) and non-essential gene (non-EG) sets from https://github.com/hart-lab/bagel, 684 EGs and 926 non-EGs). The essential set was used as the true positive list for the precision_recall_curve function of the Scikit-learn v0.23.0 library for Python, along with the above BF score subset[8]. For generating fold change plots of EGs and non-EGs, we calculated the mean fold change of all guides for a gene for each screen. Replicates were grouped together and mean fold change for each gene was calculated. The x axis was defined as [min(fold_change), max(fold_change)] on a 0.2 interval. Gaussian kernel density estimation of fold changes was performed separately for essential and non-essential training sets with kde.factor 0.25. using gaussian_kde from Scipy v1.5.2 Python library[9]. Essential and non-essential data were plotted separately for each grouped set of replicates.

Gene hit selection

To select genes that affect cell growth in the context of CTCF depletion, we performed three steps. We used R v4.1.0 to perform this analysis.

First, we compared the log2 fold changes (log2 FC), calculated to the T0 reference sample, of the HAP1-CTCFdegron screens against the HAP1-CTCFdegron-TIR1 screens for each time point (T6 and T15) and condition (NT, 25µM and 500µM auxin (IAA)) for a total of six comparisons. We then calculated the difference between the log2 FC HAP1-CTCFdegron-TIR1 with log2 FC HAP1-CTCFdegron for each condition and time point
(log2 FC HAP1-CTCFdegron-TIR1 - log2 FC NO HAP1-CTCFdegron).

We selected genes that had a log2 FC difference > |1|, i.e., a fold change > |2| for each time point and condition. We calculated the log2 FC difference separately for positive and negative values (i.e., < -1 or > 1). A positive value indicates that the cells were growing better when CTCF was depleted, and the target gene was knocked out; we call them positive interaction hits. A negative value indicates that cell growth was affected when CTCF was depleted, and the target gene was knocked out; we call them negative interaction hits. This selection resulted in hit lists for positive and negative interactions for each time point and condition. We then merged these gene lists, keeping positive or negative interaction hits separated. We next wanted to remove from this analysis the interactions caused by auxin only. Therefore, we performed the same analysis as described above for comparing NT versus 25µM or 500µM auxin only in the HAP1-CTCFdegron cells (log2 FC HAP1-CTCFdegron auxin treated (25µM or 500µM) - log 2 FC HAP1-CTCFdegron NT).

The “hits” obtained this way were labeled auxin specific genes and removed from the overall hit lists.

Second, independently of step one, we calculated the ratio of the FC auxin-treated by the FC NT (log2(FC treated with auxin (25 or 500µM auxin) / FC NT).

We compared these ratios from the HAP1-CTCFdegron screens against the HAP1-CTCFdegron-TIR1 screens for each time point (T6 and T15) and condition (25µM and 500µM auxin) for a total of four comparisons. We then calculated the difference between the ratio of HAP1-CTCFdegron-TIR1 with the ratio of HAP1-CTCFdegron for each condition and time point (ratio HAP1-CTCFdegron-TIR1 - ratio HAP1-CTCFdegron).

We selected genes that had a difference > |1|, ie a fold change > |2| for each time point and condition and merged the hit lists as described above, again keeping positive and negative interactions separate. We again removed the auxin specific genes from these lists.

Third, we calculated the union of the gene hits from step 1 and step 2, separately for positive and negative interactions. The final gene list had 469 negative interaction hits and 294 positive interaction hits.

GO analysis

Functional enrichment in the gene hit lists was calculated using Enrichr (https://maayanlab.cloud/Enrichr/), GO Biological Process 2018[10].

CTCF concentration time course from Fig. 4d

sgRNA depletion or enrichment was assessed over decreasing CTCF levels (ranging from full amount of CTCF to nearly no CTCF. 0, HAP1-CTCFdegron, full amount of CTCF; 1, HAP1-CTCFdegron-TIR1 NT, decreased level of CTCF; 2, HAP1-CTCFdegron-TIR1 auxin 25µM, further decrease of CTCF level; 3, HAP1-CTCFdegron-TIR1 auxin 500µM, nearly no CTCF

left). We calculated the log2 fold change for each gene for each CTCF concentration to its respective T0 sample. Log2 fold changes were then subtracted with the respective log2 fold changes in the HAP1-CTCFdegron samples, and the resulting matrices were plotted as heatmaps. Genes that change (indicated as red in Fig. 4d) have a difference of log2 fold change of > |0.4|. Genes with similar patterns were manually grouped together.

RNA-seq analysis

Bulk RNA-seq analysis

RNA-seq data was mapped using a standardized pipeline in the DolphinNext environment[11]. Briefly, reads were mapped to hg19 using STAR v2.6.1 with default parameters to generate BAM files[12]. RSEM v1.3.1 was used for transcript quantification[13]. To identify Differentially Expressed genes, DESeq2 v3.15 was run using the DEBrowser interface v1.20.0 (see Supplementary Table 3 for RNA-seq statistics)[14]. Genes that had less than 10 counts per million were removed using Low Count Filtering method. The dispersion was estimated using a parametric fit and hypothesis testing performed using the likelihood ratio test. We selected differentially expressed genes that were at least 1.25-fold up- and down-regulated, with p-values less than 0.05.

Splicing RNA-seq analysis

For alternative splicing (AS) analyses, reads were trimmed with Trimmomatic v0.32, aligned with STAR v2.70e and AS events detected with RMATS v4.1.0[15–17]. AS events with FDR<0.05 were considered as statistically significant AS events. DDX55 knock-out clones in absence of auxin (NT), DDX55 knock-out clones in presence of auxin (IAA), TAF5L knock-out clones in absence of auxin (NT), TAF5L knock-out clones in presence of auxin (IAA) were pooled as technical replicates for RMATS analyses.

ChIP-seq analysis

Raw data processing

50bp paired ends reads were processed using the nf-core/chipseq pipeline, version 1.1.0: https://github.com/nf-core/chipseq[18]. Briefly, fastq files were mapped to the hg19 reference genome using BWA, mapped reads were filtered to remove duplicates, unmapped reads, multi-mappers, etc, and finally bigWig files scaled to 1 million mapped reads were created. ChIP-seq samples passed QC-tests (see Supplementary Table 4 for ChIP-seq statistics).

Peak calling

We used MACS2 v2.2.7.1 to find the enriched ChIP-seq peaks with the following parameters: macs2 callpeak -q 0.01[19]. To have a final list of ChIP-seq peaks for each individual protein (CTCF, RAD21, DDX55 and TAF5L), we took the union of the peaks called in the two ChIP-

seq replicates. Peaks that were as close as 10bp or overlapping were merged together using bedtools merge -d 10 (v2.29.2). Intersection between the peaks was done using bedtools intersect v2.29.2[20].

Motif search analysis

We used MEME v5.0.5 to find motifs in sequences corresponding to ChIP-seq peaks and intersection of ChIP-seq peaks. Following parameters were used: -mod anr -nmotifs 3 -minw 15 -maxw 50 -objfun classic -revcomp -markov_order 0[21,22].

Peak annotation

Peaks were annotated using annotatePeaks.pl from HOMER v4.6[23].

Venn diagrams

Venn diagrams were generated using intervene venn v0.6.4 and bed files[24].

Aggregation stackups

Series of stackup panels were generated using custom scripts. Each panel is demonstrating behavior of a number of *signals* - in columns (e.g., ChIP-seq, insulation, EV1, RNA-seq etc.) around a set of genomic loci of interest, *features* - in rows (e.g., ChIP-seq peaks, genes TSS/TTS, subsets of the above, etc.). In a given column, *signals* are extracted using bbi.stackup function from Python API for UCSC BBI library v0.3.2 (https://github.com/nvictus/pybbi) centered on a given feature with a given flank[25]. Every column in a given stackup panel is sorted synchronously, and a summary (average) profile of every *signal* is presented in the axes above.

We used following *signal*-specific parameters for stackups (unless stated otherwise in a figure caption):

- all ChIP-seq, RNA-seq and DRIP-seq derived signals used 200 bp-sized bins with a 5kb flank, R-loops signal used 10kb flank instead.
- Insulation signal for 20kb - *diamond* used 2kb - sized bins and flank of 40kb, whereas insulation with 100kb - *diamond* used 5kb - sized bins and 200kb flanks.
- EV1 signal used 25kb - sized bins and 150kb - sized flanks.

When specified on a figure caption, signals extracted for gene-related features (TSSs/TTSs) were flipped for genes on the negative strand, to ensure that the body of a given gene is on the right side of the stackup for TSSs, and on the left side of the stackup for TTSs.

Stackups were sorted according to the signal specified in a figure caption using average of the central bins, with the exception of stackups sorted on R-loops and stackups sorted on EV1 signal: average of an entire row was used for sorting R-loops signal at TTSs, due to the wide nature of the signal; drop of EV1 signal, calculated as a difference of average EV1 between

the left and the right flank, was used for sorting stackups using EV1 - this allowed us to reveal apparent B to A and A to B transitions near the center of each feature.

Linear color scales were used throughout the stackups with the following exceptions:

- RNA-seq signal was plotted using log-spaced colormap and the corresponding average profile is a geometric mean of RNA-seq signal
- discrete signals extracted from bigBED files: R-loop used blue color to indicate footprints of peaks and white color to indicate regions without any peaks; gene annotation used red/blue colors to indicate genes on the positive/negative strands and grey color to indicate regions without annotated transcripts.

To plot the distribution of CTCF and RAD21 peaks in presence of CTCF, we calculated the rolling sum of those peaks with the window size 100 and plotted the sums for the non-overlapping windows along the stackup.

Insulation signal was subject to additional normalization before plotting on the stackup - we used average insulation signal of the outermost half-flank upstream and downstream to normalize each individual insulation profile on the stackup. This normalization was done to alleviate fluctuations of insulation signal along the chromosomes (e.g., near-telomeric regions, compartmental variation, etc).

List of features used in the paper

List of biologically relevant peaks used in Fig. 1

We have assembled a list of biologically relevant peaks, by combining CTCF and RAD21-peaks: we merged together all the CTCF peaks in presence of CTCF (NT) and all the RAD21 peaks in presence and absence of CTCF (NT and IAA). The final list contains 39,233 peaks.

List of genes (Gene annotation)

We downloaded gene annotation RefGene for hg19 human reference genome from UCSC, and matched it with gene description gene_info from NCBI using gene2refseq dictionary. Genes with identical names and TSS coordinates were merged together to create longest possible transcripts using exonU.py script from clodius v0.3.5 (https://github.com/higlass/clodius). Resulting gene annotation was used to create gene-related feature lists in the manuscript.

Active TSSs in HAP1

We used publicly available active mark H3K4me3 ChIP-seq signal to rank annotated TSSs from the most active to inactive[26,27]. Top 9,967 TSSs were selected as Active TSSs in HAP1 (H3K4me3 signal drops to nearly 0 after the top 9,967 TSSs).
We further filtered out active TSSs that were too close to a called CTCF peak (<2kb) in HAP1, in order to avoid interfering signals from CTCF-insulation and CTCF-related interactions. The final list contains 8,849 active TSSs without CTCF in HAP1.

Active TSSs common in HAP1 and HCT116 cells
First, we generated a list of Active TSSs for HCT116 cell line following the same procedure as for HAP1 cell line, using publicly available active mark H3K4me3 for HCT116 (ENCODE, HCT116). Intersection of this list with the list of active TSSs in HAP1 yielded 12,113 TSSs that are active in both cell lines. To avoid interfering CTCF insulation and interactions in the analyses, we further filtered our active TSSs that have CTCF by doing the union of all the CTCF peaks in HAP1 and HCT116 cell lines and intersecting it with the common list of active TSSs. The final list contains 6,802 common active TSSs without CTCF.

Active TTSs in HAP1 cells
We selected active TTSs by taking the list of active TSSs in HAP1 and attributing for each active TSS its TTS. To avoid having interfering TSS insulation in the TTS stackups, we removed TTSs that have H3K4me3 signal nearby. To avoid having interfering CTCF insulation in the TTS stackups, we intersected the active TSSs without H3K4me3 (extended 2kb each side) with the CTCF peaks using bioframe v0.3.1[28]. The final list contains 10,933 active TTSs without CTCF.

Active TTSs common in HAP1 and HCT116 cells
To create a common list of active TTSs without H3K4me3 between HAP1 and HCT116 cell lines, the intersection between the active TTSs without H3K4me3 in HAP1 and the active TTSs in HCT116 was done. To avoid having interfering CTCF insulation in the TTSs without H3K4me3 stackups, we removed the active TTSs without H3K4me3 that have CTCF by doing the union of all the CTCF peaks in HAP1 and HCT116 cell lines and intersecting it with the common list of active TSSs without H3K4me3. The final list contains 8,894 common active TTSs without CTCF.

Consensus list of R-loops
We created a consensus list of R-loops by intersecting different DRIP-seq datasets published in[29]. We intersected the NT2 and K562 datasets (extended 2kb each side) by using bedtools intersect v2.29.2. The final list contains 19,229 R-loops.

Enhancers

We created a list of HAP1 enhancers by overlapping H3K27ac peaks with HAP1 DNase-seq signals (ENCODE, HAP1). We filtered out enhancers that were too close to a called CTCF peak (<2kb), in order to avoid interfering signals from CTCF-insulation and CTCF-related interactions. We further filtered out TSSs that were too close to an active TSS (H3K4me3 active mark) (<2kb), in order to avoid interfering signal from TSS insulation. The final list contains 10,443 enhancers.

CTCF motif orientation

CTCF peaks were oriented using one CTCF motif in Jaspar (MA0139.1). If a CTCF peak had multiple CTCF motifs, the motif with the strongest p-value was attributed to the CTCF peak.

Hi-C analysis

Hi-C data processing

Hi-C libraries were processed using the distiller pipeline v0.3.3 (https://github.com/open2c/distiller-nf). Briefly, paired-end reads were mapped to the human reference assembly hg19 using bwa mem in a single-sided fashion (-SP). Then the pairtools package v1.0.0 (https://github.com/open2c/pairtools) was used to parse alignments and classify ligation products, or pairs. Uniquely mapped and rescued pairs were kept after removal of duplicates. Pairs were further filtered, keeping only those with high mapping quality scores on both sides (MAPQ > 30), and subsequently aggregated into binned contact matrices in the cooler format v0.8.11 at 1, 2, 5, 10, 25, 50, 100, 250, 500 and 1000kb resolutions[30]. All contact matrices were normalized using the iterative correction normalization after excluding the first 2 diagonals to avoid short-range ligation artifacts at a given resolution[31]. Bins with extreme coverage were excluded using the MADmax (maximum allowed median absolute deviation) filter on genomic coverage, described in[32], using the default parameters (see Supplementary Table 5 containing the Hi-C statistics and Supplementary Table 7 for published datasets). Sex chromosomes (chrX, chrY) and mitochondrial genome (chrM) were left out from Hi-C analysis. The data was initially inspected using HiGlass v1.11.7[33]. For the figures, Hi-C interaction heatmaps were plotted using matplotlib v3.5.2[34].

Scaling plots

We used balanced Hi-C data binned at 1kb to calculate average frequency of interactions P(s) between pairs of loci separated by the same genomic distance (s) independently for each chromosomal arm, using compute_expected from cooltools v0.5.0[35]. As Hi-C data becomes very sparse at large genomic separations, average frequencies were grouped into log-spaced

bins of genomic separation and averaged using logbin_expected from cooltools v0.5.0. These log-binned distance decay curves for each chromosomal arm were combined together to summarize distance decay genome-wide and calculate the rate at which interactions decay with distance, i.e., scaling plot derivatives, using combine_binned_expected from cooltools v0.5.0.

Average distance decay P(s) calculated using compute_expected from cooltools v0.5.0, at a given resolution was also used as an expected (to calculate observed-over-expected signal, OE) for a number of subsequent analyses including eigendecomposition, interaction pileups and dot pileups.

Compartments

A and B compartments were assigned using an eigenvector decomposition procedure implemented in the cooltools v0.5.0 package[31,35]. Eigenvector decomposition was performed on observed-over-expected cis contact matrices for every chromosome arm binned at 25kb and 100kb. The first eigenvectors (EV1) positively correlated with the gene density were used to assign A and B compartment identity to each bin.

The strength of compartmentalization was analyzed using saddle-plots as implemented in cooltools v0.5.0[31]. Briefly, EV1 were digitized into 50 quantiles (excluding 2% of outlier values) $EV1_i$ i=1,2,..,50, and observed-over-expected (OE) contact signal was averaged across chromosomal arms for every combination of quantiles $EV1_i$ - $EV1_j$.

We estimated the strength of A compartment as an enrichment of AA interactions over AB: AA / ((AB+BA)/2), where AA is an average of OE interactions between EV1 quantiles with 20% strongest A-compartment identity, and AB(=BA) is an average of OE interactions between quantiles with 20% strongest A- and B-identities. Similarly for B-compartment, strength was estimated as: BB / ((AB+BA)/2).

Insulation

Insulation scores were calculated using cooltools v0.5.0 implementation of diamond insulation method[36]. Insulation at TSSs was calculated using a diamond size of 20kb and a bin size of 2kb, unless specified otherwise in the text and/or figure caption. Insulation at TTSs was calculated using a diamond size of 100kb and a bin size of 5kb.

Contact frequency pileup heatmaps

In order to analyze the average interaction pattern at a number of selected locations (on-diagonal pileups) and for a number of selected pairwise features (off-diagonal pileups) we used snipping module from cooltools v0.5.0. Briefly, individual slices of observed-over-expected interaction heatmap centered on selected features and extended for the specified genomic distances were extracted and averaged.

Dot selection

Dots in the HAP1 cell line have been characterized in[37]. The CTCF degron was built in the HAP1 cell line however the CTCF degron is now different from the HAP1 parental cell line. As described in the result part, the CTCF degron has also a lower level of CTCF. Some dots that were present in the parental HAP1 cell line are not present in the CTCF degron. In order to only select the dots that are present in the CTCF degron cell line, we overlap the Sanborn dot anchors with the CTCF degron ChIP-seq peaks (strength > 150). We only selected the dot anchors that had a CTCF peak. We then only selected the dots that had a CTCF peak on either anchor. The final list has 4,661 dots. For the dots that had remaining CTCF after CTCF depletion, we took the CTCF degron list of dots and selected the dots that had remaining CTCF after CTCF depletion (CTCF ChIP-seq after auxin treatment) on either dot anchor. There are 1,604 potential dots that have remaining CTCF after CTCF depletion, but they are not able to form interactions.

All pairwise interaction pileups

CTCF-TSS

All pairwise combinations between CTCF (without TSSs and with RAD21 binding) and active TSSs (without CTCF) separated by 50-500kb were generated using pair_by_distance from bioframe v0.3.1.

CTCF-TSS nearest

Closest CTCF (without TSSs and with RAD21 binding) or TSSs (without CTCF) were identified using closest from bioframe v0.3.1 independently for upstream or downstream CTCF sites. The two lists were then concatenated and CTCF and TSSs separated by 50-500kb were kept.

Promoter-enhancer

All pairwise combinations between active TSSs (without TSSs and with RAD21 binding) and enhancer (without CTCF) separated by 50-500kb were generated using pair_by_distance from bioframe v0.3.1.

Gene meta-analysis

Demonstration of insulation profiles and interaction landscapes around genes was performed separately for *active* and *inactive* gene categories. We used the list of common active TSSs

in HAP1 and HCT116 to define *active* genes for the analysis on Fig. 2b, while for analyses on Fig. 6b a*ctive* genes were defined as genes with expression levels >= 5 TPM across all generated RNA-seq samples. *Inactive* genes were defined as common inactive TSSs (using exclusion of H3K4me3 signal) in HAP1 and HCT116 for the analysis on Fig. 2b; while for analyses on Fig. 6b, *inactive* genes were defined as genes with 0 TPM expression levels across all generated RNA-seq samples. We considered genes only with transcript sizes >= 50kb for all categories, as it is difficult to interpret Hi-C data for genes smaller than 10-bins in size (5kb bin size). We further excluded genes that have detectable levels of CTCF ChIP-seq signal in a 2kb vicinity of TSS or TTS, in order to avoid any insulation or interaction patterns associated with CTCF.

Metagene insulation profile analysis

Genomic intervals between TSS and TTS were extended by 100%, and then used to extract a profile of the signal (insulation and R-loop signals) from corresponding bigWig files. We used bbi.fetch with 3,000 bins to extract profiles rescaled to the same number of bins regardless of the interval size[25]. Extracted profiles were further averaged and in the case of insulation - normalized by subtracting the mean value of the outermost half of both flanks.

Rescaled pileups for genes and enhancer-promoter interactions

We used the coolpuppy v0.9.5 package (https://github.com/open2c/coolpuppy) to illustrate average interaction landscape around select genes and enhancer-promoter (EP) pairs[38]. On diagonal interaction-pileups were generated using genomic intervals between TSS and TTS of active genes without CTCF binding and genomic intervals between enhancers and TSS as EP interaction pairs.

For both types of genomic regions, rescaled pileups were generated as follows: regions of interest were extended 100% up- and down-stream (flanking) and corresponding Hi-C matrices were extracted at 5kb resolution. Extracted matrices were normalized by the randomly shifted local controls, intelligently rescaled to 200*200 pixels using zoom_array from cooltools v0.5.0, and finally averaged for the regions of interest.

Rescaled pileups were generated independently for positively- and negatively-stranded genes. Pileups for negatively stranded genes were then flipped to match the orientation of positively stranded genes and combined with the pileups of the latter ones. Similar transformation was done to combine matching pileups of promoter-enhancer regions: upstream enhancer - (positive strand)TSS combined with the downstream enhancer - (negative strand)TSS; upstream enhancer - (negative strand)TSS combined with the downstream enhancer - (positive strand)TSS, etc…

Jupyter notebooks used in this study can be found on github:
https://github.com/dekkerlab/ALV-repo.git

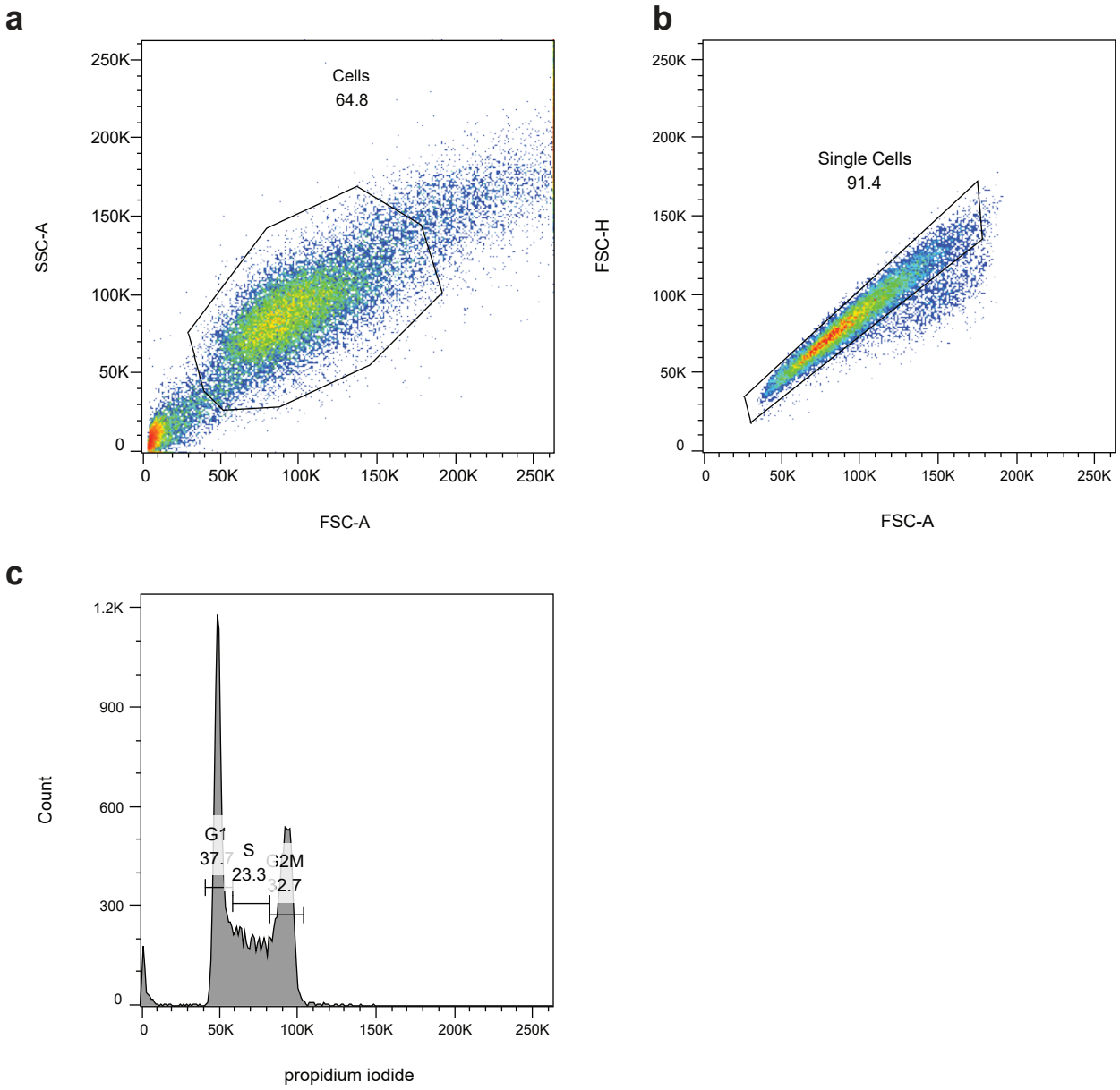
The data datasets generated in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO SuperSeries accession number GSE180691[39]. This SuperSeries is composed of the following SubSeries: GSE180922 (Hi-C), GSE180713 (RNA-seq), GSE180690 (ChIP-seq), GSE180657 (CRISPR screen).

REFERENCES

1. Hart, T. *et al.* High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **163**, 1515–1526 (2015).
2. Hart, T. *et al.* Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. *G3 Bethesda Md* **7**, 2719–2727 (2017).
3. Aregger, M., Chandrashekhar, M., Tong, A. H. Y., Chan, K. & Moffat, J. Pooled Lentiviral CRISPR-Cas9 Screens for Functional Genomics in Mammalian Cells. *Methods Mol. Biol. Clifton NJ* **1869**, 169–188 (2019).
4. Belaghzal, H., Dekker, J. & Gibcus, J. H. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods San Diego Calif* **123**, 56–65 (2017).
5. Mair, B. *et al.* Essential Gene Profiles for Human Pluripotent Stem Cells Identify Uncharacterized Genes and Substrate Dependencies. *Cell Rep.* **27**, 599-615.e12 (2019).
6. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
7. Hart, T. & Moffat, J. BAGEL: a computational framework for identifying essential genes from pooled library screens. *BMC Bioinformatics* **17**, 164 (2016).
8. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
9. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
10. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90-97 (2016).
11. Yukselen, O., Turkyilmaz, O., Ozturk, A. R., Garber, M. & Kucukural, A. DolphinNext: a distributed data processing platform for high throughput genomics. *BMC Genomics* **21**, 310 (2020).
12. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–21 (2013).
13. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
14. Kucukural, A., Yukselen, O., Ozata, D. M., Moore, M. J. & Garber, M. DEBrowser: interactive differential expression analysis and visualization tool for count data. *BMC Genomics* **20**, 6 (2019).
15. Park, J. W., Tokheim, C., Shen, S. & Xing, Y. Identifying differential alternative splicing events from RNA sequencing data using RNASeq-MATS. *Methods Mol. Biol. Clifton NJ* **1038**, 171–179 (2013).
16. Shen, S. *et al.* MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. *Nucleic Acids Res.* **40**, e61 (2012).
17. Shen, S. *et al.* rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E5593-5601 (2014).
18. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).
19. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
20. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma. Oxf. Engl.* **26**, 841–842 (2010).
21. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36 (1994).
22. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202-208 (2009).
23. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

24. Khan, A. & Mathelier, A. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics* **18**, 287 (2017).
25. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinforma. Oxf. Engl.* **26**, 2204–2207 (2010).
26. Campagne, A. *et al.* BAP1 complex promotes transcription by opposing PRC1-mediated H2A ubiquitylation. *Nat. Commun.* **10**, 348 (2019).
27. Mchaourab, Z. F., Perreault, A. A. & Venters, B. J. ChIP-seq and ChIP-exo profiling of Pol II, H2A.Z, and H3K4me3 in human K562 cells. *Sci. Data* **5**, 180030 (2018).
28. Abdennur, N. *et al. open2c/bioframe: v0.2.0*. (Zenodo, 2020). doi:10.5281/zenodo.4302679.
29. Sanz, L. A. *et al.* Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Mol. Cell* **63**, 167–178 (2016).
30. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinforma. Oxf. Engl.* **36**, 311–316 (2020).
31. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
32. Schwarzer, W. *et al.* Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
33. Kerpedjiev, P. *et al.* HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* **19**, 125 (2018).
34. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
35. Venev, S. *et al. open2c/cooltools: v0.4.0*. (Zenodo, 2021). doi:10.5281/zenodo.4667696.
36. Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).
37. Sanborn, A. L. *et al.* 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 (2015).
38. Flyamer, I. M., Illingworth, R. S. & Bickmore, W. A. Coolpup.py: versatile pile-up analysis of Hi-C data. *Bioinforma. Oxf. Engl.* **36**, 2980–2985 (2020).
39. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).

SUPPLEMENTARY FIGURE



Supplementary Figure 1 : Gating strategy for flow cytometry experiments
(a) Side scatter area plotted against forward scatter area to separate cell events from debris
(b) Forward scatter height plotted against forward scatter area to separate single cells from aggregates
(c) Histogram of propidium iodide (DNA content) to assess the proportion of cells in G1, S and G2/M phases