nature research

Corresponding author(s): Di Zhang, Gerd Blobel

Last updated by author(s): Jul 13, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	For processing and analyzing next-gen sequencing data, we used the following tools and packages: BBDuk tool 37.68 bowtie 2: 2.3.3.1 SAMtools: 1.5 BEDtools: v2.17.0 HiC-Pro_2.9 and HiC-Pro 2.11.3-beta Lib5C 0.5.3 matplotlib 2.2.3 pandas 0.22.0, scipy 0.19.1, numpy 1.13.3, juicer_tools.1.7.6 FastQC 0.11.5 trim galore: 0.4.1-0 cutadapt: 1.18 FLASH: 1.2.8 CCAnalyzer v3 Salmon v0.9.1. DESeq2 1.26.0 3D netmod domain calling (v3.0_development) as described in Norton et. al., 2018 (DOI: 10.1038/nmeth.4560) and in Zhang et. al., 2019 (DOI: 10.1038/s41586-019-1778-y) FastQC v0.10.1 STAR_2.5.1b

kentUtils UCSC Genome Browser v369
MACS 1.3.7.1
Sicer V1.1
bamCoverage 3.1.3
DiffBind v2.14.0
tximport_1.14.2
For statistical analyses and implementation of insulation score calculations, we used R (3.5.0)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All main, extended data and supplementary figures include publicly available data. All Hi-C, Capture-C, RNA-seq, ChIP-seq, and other applicable next-gen sequencing raw data and processed data generated from this study are now available as GSE137376 on GEO database. Mouse CTCF ChIP-seq and mouse Hi-C domain boundaries (both asynchronous) shown in Figure 6a-c are derived from Zhang et. al., 2019 (DOI: 10.1038/s41586-019-1778-y), the updated GEO database for which has the accession number GSE129997. In Supplementary Figure 1: Hi-C heatmaps from all cell lines, except for HAP1, are from from GEO: GSE63525 by Rao et. al., 2014 (DOI: 10.1016/j.cell.2014.11.021); K562 ChIP-seq data are from ENCODE, CTCF (DCC accession: ENCSR000AKO), SMC3 (DCC accession: ENCSR000EGW), RAD21 (DCC accession: ENCSR000FAD) and Pol2 (DCC accession: ENCSR000FAY).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was not pre-determined. We used sample sizes commonly accepted for next-gen sequencing based genome-wide experiments (Huang et. al., 2017 (DOI: 10.1101/gad.303461.117); Zhang et. al., 2019 (DOI: 10.1038/s41586-019-1778-y); Despang et. al., 2019 (DOI: 10.1038/s41588-019-0466-z)). For each genotype (wildtype or genome-edited cell line) whose Hi-C results are shown, we performed 2 independent Hi-C experiments. For each genotype (wildtype or genome-edited cell line) whose Capture-C results are shown, we performed at least 2 independent Capture-C experiments. Hi-C and Capture-C data were pooled for down-stream analyses. For each genotype (wildtype or genome-edited cell line) whose RNA-seq and CTCF/RAD21 ChIP-seq results are shown, we also performed 2 independent RNA-seq experiments and ChIP-seq of CTCF and RAD21. For each genotype whose target-enriched insertion mapping and H3K27Ac ChIP-seq results are shown, we performed 1 experiment.
Data exclusions	No data were excluded from analyses.
Replication	2 biological replicates (2 independent experiments per genotype) were performed for Hi-C, RNA-seq, and ChIP-seq of CTCF and RAD21. At least 2 biological replicates (2 independent experiments per genotype) were performed for Capture-C. Data from all replication attempts were included. Mapping and other quality metrics were compared between replicates for Hi-C and Capture-C before data from replicates were pooled. Mapping metrics were examined, and data tracks were visualized in multiple genomic regions for RNA-seq replicates prior to downstream differential expression analysis. ChIP-seq peak call concordance and data tracks were examined for each ChIP-seq replicate.
Randomization	Experiments were not randomized. No animal or human subjects were involved in this study.
Blinding	Investigators were not blinded to allocation during experiments and outcome assessment. Blind was not relevant to our study as no human subjects were involved.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Histone H3K27ac antibody (Mouse Monoclonal), Active Motif, catalog number 39685, lot numbers 33417016 and 5919019. Dilution: 10ug/ChIP.

CTCF antibody (Rabbit Polyclonal), EMD Millipore, catalog number 07-729, lot number 3273150. Dilution: 10uL stock/ChIP.

RAD21 antibody (Rabbit Polyclonal), abcam, catalog number ab992, lot numbers GR3253930-6, GR3310168-1, and GR3310168-2. Dilution: 10ug/ChIP.

Mouse IgG, Sigma, catalog number I8765, lot number SLBW2188. Dilution: 50ug for pre-clearing and 10ug/ChIP.

Rabbit IgG, Sigma, catalog number I8140, lot number SLBT3686. Dilution: 50ug for pre-clearing and 10ug/ChIP.

Validation

The Histone H3K27ac antibody has been claimed to react with human and to be ChIP-grade per manufacturer. This antibody has also been previously used in our lab (Behera et al. 2019, DOI: 10.1016/j.celrep.2019.03.057), as well as 15 other publications that used and cited this antibody on manufacturer's website.

The CTCF and RAD21 antibodies have been claimed to react with human and to be ChIP-grade per the manufacturers. These two antibodies have also been previously used in our lab (Zhang et. al. 2019, DOI: 10.1038/s41586-019-1778-y), and in other published studies.

The mouse and rabbit IgGs have been previously used in our lab for pre-clearing and for isotype-matched control (Zhang et. al. 2019, DOI: 10.1038/s41586-019-1778-y). They are commonly used as control for ChIP per the manufacturer, and have been cited in at least 17 and 25 publications listed on the manufacturer's website.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HAP1 cells are near-haploid cells derived from KBM7 cells (Carette et al., Nature 477, 340-343 (2011)).
Authentication	HAP1 cells were routinely stained with a cell-permeant dsDNA dye, followed by sorting.
Mycoplasma contamination	HAP1 cells were tested for Mycoplasma.
Commonly misidentified lines (See ICLAC register)	The cell line used (HAP1) is not in the ICLAC database.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	All Hi-C, Capture-C, RNA-seq, ChIP-seq, and other applicable next-gen sequencing raw data and processed data generated from this study are now uploaded into the GEO database, under super series GSE137376. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE137376
Files in database submission	Please see the GEO template sheet: https://upenn.box.com/s/2qqvmsmcdd5f90egoro049zdcfwtngg3
Genome browser session (e.g. UCSC)	

1. Open Integrated Genomics Viewer (IGV)
2. Download these IGV session files to local drive:
   -Fig. 4f: https://upenn.box.com/s/kpslkten6052qblo0hsig58kg0q06nda
   -Fig. 4g: https://upenn.box.com/s/1xz5i2xv1oxuuvn7p2rn3dllox599xjr
   -Fig. 4h: https://upenn.box.com/s/x1upxfdjhlyp8roww3kcnr7mhx5ph1dp
   -Fig. 4i: https://upenn.box.com/s/tklurgo5oj4qecm8uukz5krpxbs4m02k
   -Fig. 5f: https://upenn.box.com/s/s8m1ua7ed7rqg8y73l6ghozvw5cmtjod
   -Fig. 5g: https://upenn.box.com/s/knzg0sik3tohylox351ijjjs91a2qod4
   -Fig. 5h: https://upenn.box.com/s/yu4mng8qhdt62obq2451btql78arm6ol
   -Fig. 5i: https://upenn.box.com/s/idps6i0gpz0ba8pf32bcfrqg6620yfxm
3. In IGV, choose file-->open session-->select the session file downloaded from the link
4. Depending on network speed, it might take some time to fully load.
5. Alternatively, data can be downloaded to a local drive using the links in the IGV session files (.xml). Session files can be modified ("Resources" and "Tracks") for visualizing the data tracks locally.

Methodology

Replicates	As noted previously, for each genotype (wildtype or genome-edited cell line) whose Capture-C results are shown, we performed at least 2 independent Capture-C experiments. Hi-C and Capture-C data were pooled for down-stream analyses. For each genotype (wildtype or genome-edited cell line) whose RNA-seq and CTCF/RAD21 ChIP-seq results are shown, we also performed 2 independent RNA-seq experiments and ChIP-seq of CTCF and RAD21. For each genotype whose target-enriched insertion mapping and H3K27Ac ChIP-seq results are shown, we performed 1 experiment.
Sequencing depth	Hi-C for each clonal/parental populations was sequenced to generate ~248-300 million raw reads for each genotype (clonal or parental populations), with biological replicates pooled and with sub-sampling performed where necessary. This leads to ~161-173 million valid interaction pairs per genotype. Capture-C for each biological replicate was sequenced to ~5-15M raw reads per captured locus. RNA-seq for each replicate was sequenced to ~70-120M reads. H3K27Ac ChIP was sequenced to ~50-70M reads, with the input sequenced to ~46M reads. CTCF and RAD21 ChIP samples were sequenced to ~20-40M reads per replicate. Please refer to Supplementary Tables 3 and 4 for more details.
Antibodies	Histone H3K27ac antibody (Mouse Monoclonal), Active Motif, catalog number 39685, lot numbers 33417016 and 5919019. CTCF antibody (Rabbit Polyclonal), EMD Millipore, catalog number 07-729, lot number 3273150. RAD21 antibody (Rabbit Polyclonal), abcam, catalog number ab992, lot numbers GR3253930-6, GR3310168-1, and GR3310168-2.
Peak calling parameters	Basecalls using bcl2fastq2 v2.15.0.4, and parameters --barcode-mismatches 1 Mapping to reference genome hg19 canon with Bowtie 1.0.0 using parameters --chunkmbs 1024 -y -n 2 --best -k 1 --maxbts 800 -l 28 -e 80 --sam-nohead --sam Wiggle generated with MACS using parameters --nomodel --shiftsize=120--format BAM --gsize 2615371906 --tsize 36 --bw 120 --mfold 12 --wig --space 1 For H3K27Ac: Peaks called with Sicer V1.1 using parameters hg19 1 200 150 0.74 600 0.01 For CTCF and RAD21: Peaks were called with MACS using MFOLD=12, PVALUE="" --format BAM --gsize $EFFGENSIZE --tsize 36 --bw 120 --mfold $MFOLD --wig --space 1 $PVALUE Filter blacklist regions from peaks, and convert to broadpeak format (see UCSC Genome Browser for format specs)
Data quality	ChIP-qPCRs of positive control loci confirmed enrichment before the library was sent for sequencing. Raw fastq files were assessed with FastQC (v0.10.1) prior to processing. Peaks were called using input controls. Please see methods section for details with regards to Hi-C, Capture-C, and RNA-seq.
Software	We used Bowtie 0.12.8, SAMtools 0.1.18, MACS 1.3.7.1, BEDTools 2.16.2, and Sicer V1.1.