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.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

<table>
<thead>
<tr>
<th>n/a</th>
<th>Confirmed</th>
</tr>
</thead>
<tbody>
<tr>
<td>☑</td>
<td>☑</td>
</tr>
</tbody>
</table>

The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement

An indication of 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 statistics 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

Clearly defined error bars

State explicitly what error bars represent (e.g. SD, SE, CI)

Software and code

Policy information about availability of computer code

Data collection

Illumina RTA 1.18.64 (HiSeq 2500) or RTA 2.4.1 (NextSeq 500) and bcl2fastq2 v2.17 was used for basecalling and demultiplexing.

Data analysis

Data analysis was performed using R 3.4.1 and R/Bioconductor packages: TxDb.Mmusrus.UCSC.mm10.knownGene 3.4.0, rtracklayer 1.38.3, GenomicRanges 1.30.3, JASPAR2016 1.6.0, Biostrings 2.46.0, QuasR 1.18.0, Bowtie 1.18.0, edgeR 3.20.5, GOstats 2.44.0, org.Mm.eg.db 3.5.0, e1071 1.6.8, hicrep 1.4.0, HITC 1.24.0, Rcpp 0.12.18 and RcppArmadillo 0.8.5000.0.

ATAC-seq and Bis-seq reads were trimmed using cutadapt 1.15, and read alignments were performed using STAR 2.5.2b (RNA-seq), bowtie (CHIP-seq, MNase-seq, Bis-seq) or bowtie2 (ATAC-seq, HiC-seq).

HiC-seq data was aligned, filtered and normalized using HiC-Pro 2.10. TAD calling was performed using armatus 2.3.

macs2 2.1.1.20160309 was used for peak finding (ATAC-seq data), and HOMER 4.8 for motif enrichment analysis.

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 upon request. 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

The datasets generated during and analysed during the current study are available in the GEO repository under accession GSE112136 (currently private and only accessible using accession token: ajkregssljahvqr). In addition, the following public datasets were analyzed that are available in the GEO repository under the accessions GSE87819 (samples GSM2341277 to GSM2341282) and GSE94041 (samples GSM2467469 to GSM2467472).

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

- Sample size: n/a
- Data exclusions: n/a
- Replication: number of replicates indicated
- Randomization: n/a
- Blinding: n/a

Reporting for specific materials, systems and methods

Materials & experimental systems

- n/a Involved in the study
- Unique biological materials
- Antibodies
- Eukaryotic cell lines
- Palaeontology
- Animals and other organisms
- Human research participants

Methods

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

Unique biological materials

Policy information about availability of materials

Obtaining unique materials: all generated cell lines will be made available upon publication

Antibodies

- Antibodies used: Antibodies used were Snf2h (wb, AbCam, ab72499), CTCF (ChIP-seq and wb, Santa Cruz, C-20X), REST (ChIP and wb, Santa Cruz, sc-25398, H-290).
- Validation: All antibodies validated by manufacturers and previously used in several publications: Snf2h (https://www.abcam.com/SNF2H-
Eukaryotic cell lines

Policy information about cell lines

Cell line source(s) mouse ES cell line 159 was originally obtained from Miriam Bibel and is described in PMID 17546008. SNF2h ko line was generated from this line and is available upon request. Brg1 deletion line was obtained from Crabtree lab Stanford and is described in PMID 19279220.

Authentication

Genotype of cell lines was tested at the level of DNA sequence and protein (western blotting)

Mycoplasma contamination
cell lines tested negative for Mycoplasma

Commonly misidentified lines

(See ICLAC register)

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

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112132 (ChIP-seq data only)
Accession token: ajkregssljahvqr

Files in database submission

For each ChIP-seq sample, the GEO entry contains two files: the rawdata (fastq format) and a file with alignment density per 100 bp in the mouse mm10 genome (wig file):

fastq files (n=38):
ChIP_REST_SD3_1.R1.fastq.gz
ChIP_REST_SD3_2.R1.fastq.gz
ChIP_REST_SD3_input_1.R1.fastq.gz
ChIP_REST_wt_1.R1.fastq.gz
ChIP_REST_wt_2.R1.fastq.gz
ChIP_REST_wt_input_1.R1.fastq.gz
ChIP_SD3_CTCF.R1.fastq.gz
ChIP_SD3_input.R1.fastq.gz
ChIP_wt_CTCF.R1.fastq.gz
ChIP_wt_input.R1.fastq.gz
ChIP_CTCF_Brg1fl_1.R1.fastq.gz
ChIP_CTCF_Brg1fl_input_1.R1.fastq.gz
ChIP_CTCF_Brg1fl_input_2.R1.fastq.gz
ChIP_CTCF_Brg1ko_1.R1.fastq.gz
ChIP_CTCF_Brg1ko_2.R1.fastq.gz
ChIP_CTCF_Brg1ko_input_1.R1.fastq.gz
ChIP_CTCF_Brg1ko_input_2.R1.fastq.gz
ChIP_CTCF_Brg1ko_1.R1.fastq.gz
ChIP_CTCF_Brg1ko_2.R1.fastq.gz
ChIP_CTCF_Brg1ko_input_1.R1.fastq.gz
ChIP_CTCF_Brg1ko_input_2.R1.fastq.gz
ChIP_CTCF_Brg1ko_1.R1.fastq.gz
ChIP_CTCF_Brg1ko_2.R1.fastq.gz
ChIP_CTCF_Brg1ko_input_1.R1.fastq.gz
ChIP_CTCF_Brg1ko_input_2.R1.fastq.gz
ChIP_CTCF_Snf2hko-AA6_1.R1.fastq.gz
ChIP_CTCF_Snf2hko-AA6_2.R1.fastq.gz
ChIP_CTCF_Snf2hko-AA6_input_1.R1.fastq.gz
ChIP_CTCF_Snf2hko-AA6_input_2.R1.fastq.gz
ChIP_CTCF_Snf2hko-AA6_1.R1.fastq.gz
ChIP_CTCF_Snf2hko-AA6_2.R1.fastq.gz
ChIP_CTCF_Snf2hko-AA6_input_1.R1.fastq.gz
ChIP_CTCF_Snf2hko-AA6_input_2.R1.fastq.gz
ChIP_CTCF_Snf2hko-AA6_1.R1.fastq.gz
ChIP_CTCF_Snf2hko-AA6_2.R1.fastq.gz
ChIP_CTCF_Snf2hko-AA6_input_1.R1.fastq.gz
ChIP_CTCF_Snf2hko-AA6_input_2.R1.fastq.gz
ChIP_CTCF_Snf2hko-AA6_1.R1.fastq.gz
ChIP_CTCF_Snf2hko-AA6_2.R1.fastq.gz
ChIP_CTCF_Snf2hko-IP_1.scaled_0100.wig.gz

wig files (n=38):
ChIP_Snf2hko_CTCF-IP_1.scaled_0100.wig.gz
Methodology

Replicates

Between 1 and 3 biological replicates were performed per cell type and antibody (indicated).

Sequencing depth

All ChIP-seq samples were sequenced as single-end 50mers (total and uniquely mapped reads are given in Table 1 and also below, fields separated by a whitespace character):

“Sample Name” “Sample Type” “GEO Identifier” “Total Reads” “Uniquely Mapped Reads”
ChIP_Snf2hko_CTCF-IP_1 ChIP-seq GSM3058327 54200868 41000688
ChIP_Snf2hko_CTCF-IP_2 ChIP-seq GSM3058328 60425597 45493980
ChIP_Brg1fl_CTCF-IP_1 ChIP-seq GSM3331491 3122918 24765578
ChIP_Brg1fl_CTCF-IP_2 ChIP-seq GSM3331490 33897947 27018295
ChIP_Brg1ko_CTCF-IP_1 ChIP-seq GSM3331493 38507267 28856904
ChIP_Brg1ko_CTCF-IP_2 ChIP-seq GSM3331494 30647029 23010536
ChIP_addbackWt_CTCF_IP_1 ChIP-seq GSM3331485 31205439 23373905
ChIP_addbackWt_CTCF_IP_2 ChIP-seq GSM3331486 31998193 23983984
ChIP_addbackWt_REST_IP_1 ChIP-seq GSM3331483 29100199 21778609
ChIP_addbackWt_REST_IP_2 ChIP-seq GSM3331484 29252779 21922229
ChIP_Brg1fl_CTCF-IP_1 ChIP-seq GSM3331487 35381997 26855458
ChIP_Brg1fl_CTCF-IP_2 ChIP-seq GSM3331488 35361399 27018295
ChIP_Brg1ko_CTCF-IP_1 ChIP-seq GSM3331489 34920059 26511556
ChIP_Brg1ko_CTCF-IP_2 ChIP-seq GSM3331490 34183979 26161632
ChIP_Brg1fl_CTCF-input_1 ChIP-seq GSM3331481 31140107 23593001
ChIP_Brg1fl_CTCF-input_2 ChIP-seq GSM3331482 28961771 21901057
ChIP_Brg1ko_CTCF-input_1 ChIP-seq GSM3331483 29100199 21778609
ChIP_Brg1ko_CTCF-input_2 ChIP-seq GSM3331484 29252779 21922229
ChIP_Brg1fl_REST-IP_1 ChIP-seq GSM3058331 23696045 17785755
ChIP_Brg1fl_REST-IP_2 ChIP-seq GSM3058332 27028123 20288955
ChIP_Brg1ko_REST-IP_1 ChIP-seq GSM3058333 25907911 19517271
ChIP_Brg1ko_REST-IP_2 ChIP-seq GSM3058334 24504277 18414887
ChIP_Brg1fl_REST-input_1 ChIP-seq GSM3058335 18866288 14119051
ChIP_Brg1ko_REST-input_1 ChIP-seq GSM3058336 19955140 14938636
ChIP_addbackWt_REST-IP_1 ChIP-seq GSM3331467 24939210 18669645
ChIP_addbackWt_REST-IP_2 ChIP-seq GSM3331468 23637913 17771295
ChIP_Brg1fl_REST-input_2 ChIP-seq GSM3058337 25093671 19517271
ChIP_Brg1ko_REST-input_2 ChIP-seq GSM3058338 24504277 18414887
ChIP_addbackMut_REST_IP_1 ChIP-seq GSM3331469 24980361 18748328
ChIP_addbackMut_REST_IP_2 ChIP-seq GSM3331470 24192660 18236217
ChIP_addbackWt_REST_input_1 ChIP-seq GSM3331472 19676120 14667970
ChIP_addbackMut_REST_input_1 ChIP-seq GSM3331474 19678507 14723772
ChIP_Brg1fl_REST_IP_1 ChIP-seq GSM3331473 40693537 30239724
ChIP_Brg1fl_REST_IP_2 ChIP-seq GSM3331474 32043656 22906540
ChIP_Brg1ko_REST_IP_1 ChIP-seq GSM3331475 40357096 29616036
ChIP_Brg1ko_REST_IP_2 ChIP-seq GSM3331476 45424502 12285086
ChIP_Brg1fl_REST_input_1 ChIP-seq GSM3331477 38235321 28805329
ChIP_Brg1fl_REST_input_2 ChIP-seq GSM3331478 3301091 25093194
ChIP_Brg1ko_REST_input_1 ChIP-seq GSM3331479 35766209 26705309
ChIP_Brg1ko_REST_input_2 ChIP-seq GSM3331480 32481904 24249999

Antibodies

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Vendor</th>
<th>Cat. Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTCF</td>
<td>Santa Cruz</td>
<td>C-20X</td>
</tr>
<tr>
<td>REST</td>
<td>Santa Cruz</td>
<td>H-290</td>
</tr>
</tbody>
</table>

Peak calling parameters

Reads were mapped using the qAlign function from the QuasR Bioconductor package (version 1.8.0) with default parameters against BSgenome.Mmusculus.UCSC.mm10. The resulting bam files were used for peak calling. Peak calling was performed using macs2 version 2.1.1.20160309, for each cell type using a pool of all biological replicates as treatment and a pool of all input samples from all cell types as controls, with parameters: -g 1.87e9 --nomodel --shift 0 --extsize X --keep-dup all --qvalue 0.01, where X corresponds to the estimated fragment size, determined for each cell type from the cross-correlation of plus and minus strand alignment densities in promoter regions (median over all samples).

Data quality

ChiP-seq sample quality was assessed using the following criteria:
- technical quality (sequencing depth at least 12 Mio. uniquely mapped reads, unique-hit mapping rates > 70%, average read count in 1kb genomic tiles with 20% G+C not more than three-fold lower than in tiles with 60% G+C)
- reproducibility (Pearson’s r greater than 0.80 between biological replicate samples or samples of the same phenotype group, using log2 read counts in peaks regions).

Software

Illumina RTA 1.18.64 (HiSeq 2500) and bcl2fastq2 v2.17 was used for basecalling and demultiplexing. Analysis was performed using R 3.4.1 and R/Bioconductor packages: QuasR 1.18.0, Rbowtie 1.18.0. Peak calling: macs2 version 2.1.1.20160309