nature research

Corresponding author(s): Aristotelis Tsirigos, Iannis Aifantis

Last updated by author(s): Nov 2, 2019

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	All NCBI GEO datasets listed in Supplementary Table 2 were downloaded using sra toolkit version 2.8.0. RNA-Seq data from Chen et al. (see Supplementary Table 2) was downloaded via FTP from data owner upon request.
Data analysis	bowtie2 version 2.3.1. Hi-C bench. genomic-tools. R version 3.3.0. ICE-normaliztion according to Imakeav et al.. TAD calling by hic-ratio. MACS2 version 2.0.1. bedtools version 2.27.1. diffBind version 2.2.12. IGV version 2.3.83. PWMScan. deeptools version 2.3.3. STAR-aligner version 2.5.0c. ngsutils version 0.5.7. edgeR version 3.14.0. bowtie version 1.0.0. ROSE version 2015. picard-tools version XX. HiCnv.

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 sequencing data created within this study was uploaded to NCBI GEO (https://www.ncbi.nlm.nih.gov/geo/) and is available under the accession GSE115896.

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	No prior sample size determination was conducted. All experiments were conducted in at least 2 biological replicates. Statistical testing ensured significant findings.
Data exclusions	No replicates were excluded, and all attempts to replicate were successful.
Replication	All experiments were conducted in at least 2 biological replicates. For all sequencing data-types, successful replication has been confirmed with Principal Component Analysis.
Randomization	Randomization was relevant to the study, because the difference between healthy and disease was assessed.
Blinding	The investigators were not blinded to sample group allocation, because the difference between healthy and disease was assessed. Sample group assignments were further ensured using Principal Component Analysis on all relevant sequencing data.

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	CTCF (D31H2; Cell Signaling Catalog no: 3418; lot 3 &4); H3K27ac (Active motif; Catalog no: 39133, Lot no: 01518010); c-MYC (D84C12; Cell Signaling; Catalog no: 5605, dilution 1:500) Lot no: 15; Actin (Millipore, clone C4, Catalog no: MAB1501R, Lot no: 2819194, dilution 1:3000)
Validation	Validation of antibodies is ensured by commercial manufacture for the application used. For CTCF antibody from Cell Signaling, the datasheet for validation is available at https://media.cellsignal.com/pdf/3418.pdf For H3K27ac antibody from Active motif, the datasheet for validation is available at https://www.activemotif.com/documents/tds/39133.pdf For, C-MYC, the validation is available in https://media.cellsignal.com/pdf/14819.pdf

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	The CUTLL1 and Jurkat cell lines were a gift from Adolfo Ferrando's lab at Columbia.
Authentication	Cell lines have been authenticated by PCR detection of originally described translocations, detection of intra-nuclear NOTCH1 and sensitivity to originally described drugs
Mycoplasma contamination	Cell lines were tested negative for mycoplasma.

Commonly misidentified lines (See ICLAC register)	The cell lines used in this study are not listed on the ICLAC list of commonly misidentified cell lines.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	NOD-SCID-Il2rg−/− (NSG) mice between 4 to 8 weeks age
Wild animals	This study did not include wild animals.
Field-collected samples	This study did not include field-collected samples.
Ethics oversight	All animal experiments were performed in accordance with protocols approved by the New York University Institutional Animal Care and Use Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Additional co-variates (age, gender, diagnosis, mutational status) are listed in Supplementary Information.
Recruitment	Healthy T cells have been ordered commercially. Leukemia samples have been selected for two specific sub-types but potential biases are discussed in Figure 1 and Supplementary Information combining mutation status, expression and chromatin interaction information.
Ethics oversight	Samples were collected by Columbia Presbyterian Hospital or Weill Cornell Medical College with informed consent and approved and analyzed under the supervision of the Columbia University Medical Center Institutional Review Board or Weill Cornell Medical College Institutional Review Board.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

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.*	ChIP-Seq data was deposited at NCBI GEO under accession GSE115893. Token: ujmpiasozxstlwx https://urldefense.proofpoint.com/v2/url?u=https-3A__www.ncbi.nlm.nih.gov_geo_query_acc.cgi-3Facc-3DGSE115893&d=DwIBAg&c=j5oPpO0eBH1iio48DtsedbOBGmuw5jHLjgvtN2r4ehE&r=N0slyiNu-4hMYPy11ZY72eU4uF0KhiHRWO11zlklvq8&m=Iy2Owf8yNQMY1mdVP20o-vxkH3kmKtebTIULDAOM4kI&s=rAFpbkcjssrbSbe1AcbHad26xqxTtHhbyUuhv5reIf8&e=
Files in database submission	T cell CTCF rep1 T cell CTCF rep2 T-ALL1 CTCF rep1 T-ALL1 CTCF rep2 T-ALL2 CTCF rep1 T-ALL2 input Jurkat CTCF rep1 Jurkat CTCF rep2 Jurkat input CUTLL1 CTCF rep1 CUTLL1 CTCF rep2 CUTLL1 CTCF rep3 CUTLL1 CTCF rep4 CUTLL1 CTCF rep5 CUTLL1 H3K27ac rep1 CUTLL1 H3K27ac rep2 CUTLL1 H3K27ac rep3 CUTLL1 H3K27ac rep4 CUTLL1 gSI H3K27ac rep1 CUTLL1 gSI H3K27ac rep2 CUTLL1 THZ1 H3K27ac rep1 CUTLL1 THZ1 H3K27ac rep2

Genome browser session
(e.g. UCSC)

Bigwig tracks available through GEO.

Methodology

Replicates

CTCF:
T cells 2 replicates + 1 input
T-ALL 1 as 2 replicates + 1 input
T-ALL 2 as 1 replicates + 1 input
CUTLL1 DMSO as 5 replicates + 1 input
CUTLL1 gSI as 3 replicates + 1 input
Jurkat as 2 replicates + 1 input

H3K27ac:
CUTLL1 DMSO as 2 replicates + 1 input
CUTLL1 gSI as 2 replicates + 1 input
CUTLL1 THZ1 as 2 replicates + 1 input

Sequencing depth

Sequencing depth is detailed in Supplementary Table 1.

Antibodies

CTCF (D31H2; Catalog no: 3418, lot 3 and 4); Lot no:; 10 ug antibody used per IP
H3K27ac (Active motif; Catalog no: 39133) ; Lot no: Lot no: 01518010, 5ug antibody used per IP

Peak calling parameters

MACS2 parameters for CTCF: --nomodel --extsize=200 --qvalue 0.05
MACS2 parameters for H3K27ac: --broad --nomodel --extsize=200 --qvalue 0.05 --broad-cutoff 0.05

using -c option to specify input samples

Data quality

Based on merged peaks and peak strength, we performed Principal Component Analysis to ensure replication.
Total number of peaks detected with above peak-calling approach:
T cells CTCF: 34443
T-ALL 1 CTCF: 28730
T-ALL 2 CTCF: 64059
CUTLL1 CTCF: 25213
CUTLL1 gSI CTCF: 18111
Jurkat CTCF: 15196
CUTLL1 H3K27ac: 30726
CUTLL1 gSI H3K27ac: 25309
CUTLL1 THZ1 H3K27ac: 30542

Software

Read alignment: Reads were aligned against the reference sequence hg19 with bowtie2 (version 2.3.1) with standard parameters and only uniquely mapped reads were kept with MAPQ > 20.
Deduplication: Aligned reads were filtered for duplicated reads using picard-tools version 2.6.0.
Peak-calling: Peak calling for CTCF and H3K27ac was performed using MACS2 (version 2.0.1) using narrow (CTCF) and broad (--broad; H3K27ac) option (sepcial parameters: --no-model).
Differential binding: To identify differentially bound peaks, we performed diffBind (version 2.2.12) analysis, using the normalization option DBA_EDGER.
Bigwig: For visualization purposes, we generated fold-enrichment bigwig files by applying MACS2 (version 2.0.1) bdgcmp over input (-m FE)