Corresponding author(s): Khelifa Arab
Ingrid Grummt
Christof Niehrs

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).

n/a	Confirmed
☐	☒ 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)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Illumina NextSeq Control Software 2.1.0.31 Illumina RTA 2.4.11 & bcl2fastq 2.18 BD FACS Diva Software v.8.0 Epityper software v.1.2 Agilent MassHunter Workstation Software v. B.06.00
Data analysis	GraphPad Prism 7.0 Microsoft Excel 2010 Adobe Illustrator CS4 Adobe Photoshop CS4 FastQC 0.11.5 Bowtie 2.2.8 SAMtools 1.3.1 Picard 2.6.0 deepTools 2.4.3 & 2.5.1 MACS 2.1.1 BEDTools 2.25.0 SeqMonk 1.38.2 HOMER 4.9

R 3.4.1 with DESeq2 1.16.1 & IHW 1.4.0
UCSC Genome Browser & Tools

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

ChIP-Seq data are deposited in the Gene Expression Omnibus repository under accession number GSE104067

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	Sample size for experiments is indicated in the individual experiments and typically was n=3.
Data exclusions	Occasionally sample loss, experimental errors, or reagent deterioration resulted in inconclusive results. On such occasions, the entire experiment was discarded and repeated.
Replication	Experiments were replicated in independent experiments.
Randomization	No randomization was performed.
Blinding	No blinding was performed.

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

Antibodies

Antibodies used

Anti-HA rabbit polyclonal antibody (Reference : Ab-9110, company: Abcam).
Anti-FLAG M2 mouse monoclonal antibody (Reference : F-7425, company: Sigma).
Anti-TET1 (Reference: GTX-124207, company: Genetex).
Anti-DNMT3A (Reference: Ab-2850, company: Abcam).
Anti-DNMT3B (Reference: NB-300-516, company: Novus Biologicals).
Anti-GADD45A rabbit polyclonal (Reference: sc-797, company: Santa Cruz).
S9.6 antibody was purified from the hybridoma cell line HB-8739.

Validation	Anti-HA rabbit polyclonal antibody was validated in our study by ChIP-seq. Anti-DNMT3A and anti-DNMT3B antibodies were validated in ChIP experiments. Anti-GADD45A rabbit polyclonal (sc-797, Santa Cruz) was validated in Arab et al. 2014 (DOI: 10.1016/j.molcel.2014.06.031) and in the presented western blotting and ChIP experiments. S9.6 is a monoclonal antibody widely used in DRIP-seq to detect R-loops genome-wide.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293TARIDwt, HEK293TARIDmut and H387 were described in Arab et al 2014 (DOI: 10.1016/j.molcel.2014.06.031). Primary skin fibroblasts are from PromoCell (Heidelberg, Germany). The mouse embryonic stem cells harboring endogeously HA-tagged Tet1 is generated from mES cells E14tg2a, source American Type Culture Collection (ATCC), using the CRISPR-CAS9 system. HEK293 and HEK293T cells are from ATCC.
Authentication	HEK293, HEK293T, and PSFs were autenticated by vendors (ATCC, PromoCell). ES cells showed expected pluripotency characteristics. H387 cells were not authenticated.
Mycoplasma contamination	HEK293 HEK293T were tested Mycoplasma-negative. PSFs and H387 were not tested. (FLAG-HA-TET1) E14tg2a mESCs were Mycoplasma-positive.
Commonly misidentified lines (See ICLAC register)	none

Animals and other organisms

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

Laboratory animals	Three mouse embryos at stage E15.5 were collected from 3 mothers of C57BL/6 strain, the sex of the embryos was not determined.
Wild animals	This study does not involve wild animals.
Field-collected samples	This study does not involve filed-collected samples.

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.*	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104067
Files in database submission	GSM2788886_Input_Mock.mm9_track.bw GSM2788887_Input_RNaseH1.mm9_track.bw GSM2788888_Tet1_ChIP_Mock_rep1.mm9_peaks.bed.gz GSM2788888_Tet1_ChIP_Mock_rep1.mm9_track.bw GSM2788889_Tet1_ChIP_Mock_rep2.mm9_peaks.bed.gz GSM2788889_Tet1_ChIP_Mock_rep2.mm9_track.bw GSM2788890_Tet1_ChIP_Mock_rep3.mm9_peaks.bed.gz GSM2788890_Tet1_ChIP_Mock_rep3.mm9_track.bw GSM2788891_Tet1_ChIP_RNaseH1_rep1.mm9_peaks.bed.gz GSM2788891_Tet1_ChIP_RNaseH1_rep1.mm9_track.bw GSM2788892_Tet1_ChIP_RNaseH1_rep2.mm9_peaks.bed.gz GSM2788892_Tet1_ChIP_RNaseH1_rep2.mm9_track.bw GSM2788893_Tet1_ChIP_RNaseH1_rep3.mm9_peaks.bed.gz GSM2788893_Tet1_ChIP_RNaseH1_rep3.mm9_track.bw
Genome browser session (e.g. UCSC)	http://genome-euro.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=Edin&hgS_otherUserSessionName=Tet1_ChIPseq_mm9

Methodology

Replicates	ChIP-seq experiments were carried out in three biological replicates using a single mESC clone harboring HA-tagged TET1 (clone #4, Supplementary Fig. 5b).
Sequencing depth	Between 43 and 66 million paired-end 42nt (PE42) pass-filter reads per sample were obtained on NextSeq500 sequencer. More than 15 million uniquely-mapped (MapQ>=3) and duplicate-filtered concordantly mapped read pairs per sample were

	used for ChIP vs. Input peak calling and browser track generation.
Antibodies	Anti-HA rabbit polyclonal antibody (reference : Ab-9110; Company: Abcam; Lot number: GR304617-8) was used to perform ChIP-seq experiments of TET1.
Peak calling parameters	Peak calling for each of the 3 Mock-treated and 3 RNaseH1-treated ChIP samples against the respective pooled Input control was carried out with MACS v.2.1.1 and options 'macs2 callpeak -m 5 50 -g mm --bw 200 --format BAMPE'.
Data quality	Raw sequence data was quality checked with FastQC v.0.11.5. Mapped data in BAM and bigWig format was quality assessed with SAMTools v.1.3.1, PICARD v.2.6.0, deepTools v.2.4.3 and by visual inspection on the UCSC genome browser. MACS2 peak-calling produced 90482 consensus Tet1 peaks (found in at least 2 of the ChIP samples) at the default cutoff of 5% FDR.
Software	Sample demultiplexing and FastQ file generation was performed using Illumina's bcl2fastq conversion software v.2.18. The reads were aligned to the mouse reference genome (NCBIM37/mm9 build from Illumina iGenomes) using Bowtie v.2.2.8 with options '-q --phred33 --very-sensitive -p 4 --fr --end-to-end --maxins 1000'. Multi-mapped and discordant reads were removed with SamTools v.1.3.1 ('samtools view -f 2 -q 3'), and duplicated read pairs (likely PCR artefacts) were deduplicated using Picard MarkDuplicates v.2.6.0. The filtered BAM files were converted into RPKM-normalised bigWig coverage tracks using deepTools bamCoverage v.2.4.3 and parameters '--smoothLength 60 --binSize 20 --normalizeUsingRPKM --outFileFormat bigwig'. Metagene plots of average gene body read coverage were made with ngs.plot v.2.61. Principal component analysis (PCA) of sample similarity was carried out with deepTools v.2.4.3. Peaks overlapping abnormal genomic regions as determined by the ENCODE project were removed from further analysis using BedTools v.2.25.0. Downstream data analysis and visualisation was performed with SeqMonk v.1.38.2, deepTools v.2.4.3/2.5.1, HOMER v.4.9, R 3.4.1 with DESeq2 1.16.1 & IHW 1.4.0, UCSC Genome Browser with Table Browser and LiftOver tool.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Primary skin fibroblasts, HEK293TARIDwt, HEK293TARIDmut and H387 cells were used for cell cycle life sorting. Cells were stained with 5 micro M Hoechst 33342 (Sigma) for 1 h at 37°C and life sorted to G1-, S- and G2M-phase cell populations.
Instrument	BD-FACSAria II (BD Bioscience) cell sorter
Software	BD FACSDiva software
Cell population abundance	60,000 sorted cells for each experiment
Gating strategy	The gating strategy depend on the DNA content of the cells, sorted by windows leaving cells that are at the boundary out.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.