GADD45A binds R-loops and recruits TET1 to CpG island promoters

Khelifa Arab1,2,3*, Emil Karaulanov1, Michael Musheev1, Philipp Trnka1, Andrea Schäfer1, Ingrid Grummt2,4* and Christof Niehrs1,3,4*

R-loops are DNA–RNA hybrids enriched at CpG islands (CGIs) that can regulate chromatin states1–6. How R-loops are recognized and interpreted by specific epigenetic readers is unknown. Here we show that GADD45A (growth arrest and DNA damage protein 45A) binds directly to R-loops and mediates local DNA demethylation by recruiting TET1 (ten-eleven translocation 1). Studying the tumor suppressor TCF21 (ref. 7), we find that antisense long noncoding (lncRNA) TARID (TCF21 antisense RNA inducing promoter demethylation) forms an R-loop at the TCF21 promoter. Binding of GADD45A to the R-loop triggers local DNA demethylation and TCF21 expression. TARID transcription, R-loop formation, DNA demethylation, and TCF21 expression proceed sequentially during the cell cycle. Oxidized DNA demethylation intermediates are enriched at genomic R-loops and their levels increase upon RNase H1 depletion. Genomic profiling in embryonic stem cells identifies thousands of R-loop-dependent TET1 binding sites at CGIs. We propose that GADD45A is an epigenetic R-loop reader that recruits the demethylation machinery to promoter CGIs.

We showed previously that the long noncoding (lncRNA) TARID (TCF21 antisense RNA inducing promoter demethylation) recruits TET (ten-eleven translocation) enzymes to the tumor suppressor gene TCF21, leading to hypomethylation of the CpG island (CGI) promoter and increased TCF21 expression7. TARID induces TCF21 demethylation by recruiting GADD45A, which targets TET and its cofactor thymine DNA glycosylase (TDG) to specific genomic sites for DNA demethylation8–10. TCF21 is transcribed in antisense orientation to TCF21, overlapping with a CGI around the transcription start site (TSS) of TCF21. There is a GC skew downstream of the TSS of TCF21 within exon 2 of TCF21 that could favor R-loop formation (Supplementary Fig. 1a). To assay for R-loops at the TSS of TCF21, we performed DNA–RNA immunoprecipitation (DRIP) assays using the antibody S9.6, which recognizes R-loops. DRIP-qPCR (quantitative PCR) revealed R-loops at the 5’ end of TCF21 (amplicon 2), which coincides with the TSS-proximal GC skew (Fig. 1a). R-loops were observed in TARID-expressing human primary skin fibroblasts (PSFs) and in HEK293TARIDwt cells harboring an artificial TCF21-containing chromosome. No R-loops were observed in HEK293TARIDmut cells lacking the TARID promoter or in H387 cancer cells, where TCF21 is silenced by promoter hypermethylation (Fig. 1a and Supplementary Fig. 1b). Overexpression of RNase H1 (RNH1), which degrades RNA within DNA–RNA hybrids, reduced the level of R-loops at TCF21 and at RPL13A (positive control). Overexpression of RNH1 also reduced the level of TARID, TCF21 and RPL13A mRNA in PSFs and HEK293TARIDmut cells. In contrast, overexpression of a catalytically inactive but binding-competent RNH1 mutant did not reduce RNA levels (Fig. 1b and Supplementary Fig. 1c). The correlation between TARID transcription, R-loop formation and TCF21 expression supports that TARID-mediated activation of TCF21 transcription involves R-loop formation at the 5’ end of TCF21.

Next, we performed R-loop footprinting14, which relies on the ability of sodium bisulfite to mutate cytosine in single-stranded DNA to uracil (C>U>T). Amplification of bisulfite-treated DNA revealed C-to-T conversions in the region covering nucleotides −7 to +62 (relative to TSS), demonstrating that this 69-base G-rich sequence is single-stranded (Fig. 1d and Supplementary Fig. 1d). This region overlaps the CGI promoter of TCF21, which is hypermethylated in TARID-deficient tumor cells and is demethylated in cells expressing TARID15. To functionally link TARID levels to R-loop formation, we transfected H387 cells with synthetic TARID derivatives and monitored R-loops at the TCF21 promoter (amplicom 2) by DRIP (Fig. 1e). R-loop levels increased robustly after transfection with TARID full-length RNA, weakly with TARID exon 2 RNA, and not with TARID intron 1 RNA. None of the RNAs affected R-loops at RPL13A.

To link R-loop formation with DNA demethylation, we analysed methylation of the TCF21 promoter in cells transfected with TARID. Ectopic TARID induced DNA demethylation of TCF21, the demethylated region overlapping with the R-loop forming sequence −7 to +62 (Fig. 1f). Shorter TARID derivatives did not affect TCF21 methylation, demonstrating that TARID-dependent R-loop formation marks the region to be demethylated.

The concurrence of TARID-dependent R-loop formation with demethylation of TCF21 coincides with binding of GADD45A, a stress response protein that promotes active DNA demethylation16,17. We reasoned that GADD45A might directly interact with R-loops, thereby targeting the demethylation machinery to the TCF21 promoter. Indeed, chromatin immunoprecipitation (ChIP)–qPCR revealed that GADD45A occupancy was restricted to amplicom 2 comprising the R-loop (Fig. 2a). Overexpression of RNH1 reduced

1Institute of Molecular Biology (IMB), Mainz, Germany. 2Division of Molecular Biology of the Cell II, German Cancer Research Center and DKFZ-ZMBH Alliance, Heidelberg, Germany. 3Division of Molecular Embryology, German Cancer Research Center and DKFZ-ZMBH Alliance, Heidelberg, Germany.
4These authors contributed equally: Khelifa Arab, Ingrid Grummt, Christof Niehrs. *e-mail: k.arab@imb-mainz.de; i.grummt@dkfz.de; c.niehrs@imb-mainz.de

NATURE GENETICS | VOL 51 | FEBRUARY 2019 | 217–223 | www.nature.com/naturegenetics
GADD45A binding, emphasizing that association of GADD45A with TCF21 requires TARID-dependent R-loop formation.

To demonstrate GADD45A binding to R-loops, we generated a synthetic R-loop probe comprising the 5′ end of TCF21 in which either the RNA or the DNA oligonucleotide was radiolabeled (Supplementary Fig. 2a). In pull-down experiments, neither GFP nor the control RNA-binding protein PTB bound to the labeled probe, while GADD45A efficiently retained the synthetic R-loop. The interaction decreased by treatment with RNH1 but not with RNase A (Fig. 2b). Binding of GADD45A was also observed with unrelated synthetic R-loop probes, and indicated that GADD45A recognizes the hybrid structure rather than the TCF21 sequence (Supplementary Figs. 2b,c). In electrophoretic mobility assays (EMSAs) the mobility of both R-loops and DNA–RNA hybrids was reduced by GADD45A, whereas no binding to the labeled RNA oligonucleotide was observed (Fig. 2c and Supplementary Fig. 1d).
Fig. 2 | GADD45A binds R-loop structures in vitro and in vivo. a, ChIP-qPCR of GADD45A at the TCF21 promoter in PSFs overexpressing RNH1 or GFP. Data are normalized to control IgGs. Mean ± s.d., n = 3 biological replicates, two tailed t-test, *P < 0.05. b, Pull-down assay showing GADD45A binding to R-loops in vitro. Bead-bound S9.6 antibody or GADD45A (GA45a), PTB or GFP were incubated with radiolabeled R-loops followed by RNase A or RNH1 treatment. Bound probes were visualized by PhosphorImaging. c, EMSA showing binding of GADD45A to R-loops. A radiolabeled R-loop probe (upper panel) or RNA oligonucleotide (lower panel) was incubated with increasing amounts of GADD45A, followed by PAGE and PhosphorImaging. d, Competitive EMSA showing concentration-dependent displacement of GADD45A from R-loops by DNA–RNA hybrids (upper panel) but not by double-stranded DNA (dsDNA; lower panel). e, Quantification of competitive EMSAs using a GADD45A–R-loop complex and the indicated unlabeled competitors. GADD45A binding to R-loops was set to 1. ssDNA, single-stranded DNA. f, Northwestern blot showing binding of membrane-bound GADD45A to R-loops in vitro. Where indicated, the membrane was treated with RNase A or RNase A plus RNase H1. The asterisk marks a nonspecific band in the mock sample. g, GADD45A and TET1 are associated with cellular R-loops. Lysates from PSFs expressing FLAG-GADD45A (GA45a upper panel) or HA-TET1 and HA-CBP (lower panel) were immunoprecipitated with S9.6 antibody and co-precipitated proteins were visualized on Western blots. Where indicated, lysates were treated with RNH1 before immunoblotting. b–g, Experiments were repeated twice with similar results. All blot images were cropped (see Supplementary Figs. 9b–d and 10a,b).

In accordance with this, in competitive EMSA assays, single-stranded or double-stranded DNA or RNA did not attenuate binding of GADD45A to the R-loop probe, whereas DNA–RNA hybrids and R-loops efficiently competed for binding (Fig. 2d,e and Supplementary Fig. 2e). Taken together, these results show that GADD45A preferentially binds to hybrids and R-loops.

To demonstrate that the association of GADD45A with R-loops is direct rather than through associated proteins, we assayed binding of GADD45A on northwestern blots. FLAG-GADD45A as well as PTB and GFP used as positive and negative controls, were separated by gel electrophoresis, transferred to a membrane, and binding to radiolabeled probes was monitored by PhosphorImaging. GADD45A bound the R-loop probe but not the labeled RNA moiety used for R-loop formation, and binding was abolished by RNH1 treatment (Fig. 2f). Previous studies have shown that GADD45A interacts with TET1 (ref. 14), suggesting that GADD45A may target...
TET1 to cellular R-loops. Indeed, both GADD45A and TET1 were co-immunoprecipitated with the S9.6 antibody, as precipitation was abolished after RNH1 treatment (Fig. 2g and Supplementary Fig. 2f).

Simultaneous transcription of TARID and TCF21 from the same allele in both directions would lead to a head-to-head collision of RNA polymerases. A solution to this conundrum would be temporal separation of sense and antisense transcription. To examine this possibility, we monitored TARID and TCF21 transcripts in fluorescence-activated cell sorting (FACS)-sorted PSFs and HEK293TARIDwt cells. Interestingly, the levels of TARID and TCF21 mRNA oscillated with the cell cycle. TARID reached maximal levels in S-phase, whereas TCF21 mRNA was highest in G2/M-phase cells (Fig. 3a and

![Fig. 3](image-url)

Fig. 3 | Transcription of TARID precedes R-loop formation and TCF21 activation. a, RT-qPCR analysis of TARID and TCF21 mRNA in FACS-sorted PSFs. RNA levels were normalized to HPRT1 mRNA. b, DNA methylation at the TSS of TCF21 measured by MassARRAY in FACS-sorted cells. Methylation is plotted as mean value of all informative CpG residues in the amplicon scaled from 0 to 1. c, TARID and TCF21 mRNA levels in PSFs arrested at the G1/S boundary and released into the cell cycle by thymidine removal. RNA levels are normalized to HPRT1 mRNA. d, DRIP-qPCR in synchronized PSFs as in c. e, ChIP-qPCR monitoring of GADD45A occupancy at the TCF21 promoter in synchronized PSFs. The levels are normalized to control IgGs. f, RT-qPCR monitoring of TET1 mRNA in synchronized PSFs. RNA levels were normalized to HPRT1 mRNA. g, Analysis of 5hmC-modified TCF21 promoter in synchronized PSFs. 5hmC levels at CpG 7 of the TCF21 promoter (see Supplementary Fig 3b) measured by quantitative hydroxymethylcytosine analysis. h, ChIP-qPCR monitoring of HA-TET1 occupancy (left) and relative levels of 5hmC at the TCF21 promoter (right) in FACS-sorted PSFs expressing HA-TET1. i, Analysis of 5hmC at the TCF21 promoter in synchronized PSFs harvested 12 h following transfection with RNH1 or GFP. j, Schematic of peaks of TARID expression, R-loop formation, and TCF21 expression proceeding sequentially during cell cycle progression. a–i, mean ± s.d., n = 3 biological replicates. h, i, two tailed t-test. *P < 0.05, **P < 0.01, *** P < 0.001.
Fig. 4 | R-loops recruit TET1 to CGI promoters. a, LC/MS-MS analysis of genomic levels of modified cytosines after siRNA-mediated knockdown of RNH1 in control or HEK293T cells expressing HA-TET1. b, LC/MS-MS analysis of genomic levels of modified cytosines in inputs or S9.6 enriched R-loop fractions from abdomen (Abn) and head (Hd) of three mouse embryos at embryonic day 15.5. c, Differential enrichment TET1 ChIP-seq analysis of RNH1- versus mock-treated mESC samples (n = 3 biological replicates). The MA plot depicts 90,482 consensus TET1 peaks, RNH1-decreased peaks (red), and RNH1-increased peaks (blue) (FDR < 0.01). d, Metagene profiles of average TET1 ChIP-seq read coverage over all protein-coding genes. e, Peak overlap of all versus RNH1-sensitive TET1 peaks with CGI, TSS, and published R-loops5. f, UCSC browser screenshot of one RNH1-sensitive TET1 peak associated with Gadd45a TSS and a TARID-like lncRNA (E230016M11Rik). Other tracks show CGI, R-loops5, and CpG methylation. A broader view of this locus with all ChIP-seq samples is shown in Supplementary Fig. 6e. g, ChIP-qPCR monitoring of GADD45A occupancy at RNH1-sensitive TET1 peaks in the absence (Mock) or presence of RNH1 along with negative controls. Data are normalized to control IgGs. h, ChIP-qPCR monitoring of TET1 binding in human PSFs at promoters of human orthologs of the genes shown in g upon treatment with GADD45A- and GADD45B-specific siRNA or scrambled control siRNAs. a, b, mean ± s.d., n = 3 biological replicates, two tailed t-test. g, h, mean ± s.d., n = 4 biological replicates, two tailed t-test. *P < 0.05, **P < 0.01, ***P < 0.001 or indicated as not significant (n.s.).
Importantly, methylation of the TCF21 promoter also oscillated, inversely correlating with TARID levels. In PSFs, the TCF21 promoter was hypermethylated during G₁- and G₂/M-phase and hypomethylated in S-phase (Fig. 3b and Supplementary Fig. 3c). The promoter was almost fully demethylated in TARID-overexpressing HEK293TARIDkd cells and in PSFs, whereas in TARID-deficient H387 and HEK293TARIDmt cells TCF21 was hypermethylated throughout the cell cycle (Supplementary Fig. 3c).

To analyze the dynamics of TARID-dependent DNA demethylation, we monitored TARID and TCF21 mRNA levels in PSFs that were arrested at the G₁/S boundary and released into the cell cycle for different times (Supplementary Fig. 3d). While TARID was transcribed in early or mid S-phase, TCF21 mRNA was synthesized later, with transcription starting in late S-phase (Fig. 3c). This sequential transcription of antisense and sense RNA was also observed in cells released from nocodazole-arrested G₂/M cells, corroborating the idea that antisense transcription precedes sense transcription (Supplementary Fig. 4a). Parallel DRIP experiments revealed that R-loops accumulated at the TCF21 promoter in mid S-phase; that is, between maximal TARID and TCF21 expression (Fig. 3d). Only background DRIP signals were detected in any cell cycle phase when probing a control region (Supplementary Fig. 4b). In support of GADD45A binding to R-loops, GADD45A occupancy at the TCF21 promoter coincided with the DRIP signal (Fig. 3c). Moreover, the level of TET1 mRNA was markedly increased at mid S-phase, linking TARID-dependent R-loop formation to TCF21 demethylation (Fig. 3f and Supplementary Fig. 3c). Strikingly, R-loop formation and GADD45A recruitment coincided with the presence of 5-hydroxymethylcytosine (5hmC), the first oxidation intermediate of 5-methylcytosine (5mC), at the TCF21 promoter (Fig. 3g). Consistent with this, both TET1 occupancy and 5hmC levels at the TCF21 promoter were elevated in S-phase (Fig. 3h). Conversely, RNH1 overexpression reduced 5hmC levels in PSFs (Supplementary Fig. 8a). Short interfering RNA (siRNA)-mediated knockdown of GADD45A and GADD45B, which also promotes DNA demethylation, decreased binding of TET1 to these target promoters (on average 40%), supporting the idea that GADD45A promotes TET1 recruitment to R-loop-containing promoters but not to control promoters (Fig. 4h and Supplementary Fig. 8b).

In summary, our study shows that GADD45A is an epigenetic reader of promoter-associated R-loops. GADD45A may also target other factors to R-loops, such as TDG. Indeed, almost 100% of TET1 peaks overlap TDG peaks at TSS CGIs (Supplementary Fig. 8c). Although GADD45A binds to generic DNA–RNA hybrids in vitro, recruitment of TET1 by GADD45A is restricted to a small fraction of genomic R-loops in vivo. One explanation for this selectivity may be that regulatory R-loops formed by lncRNAs are molecularly distinct from R-loops formed by nascent transcripts. Although accumulation of R-loops is associated with genomic instability, R-loops generated by lncRNAs may have a beneficial role in cell homeostasis, serving as sequence-specific barcodes for GADD45A–TET1 targeting.

**Online content**

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41588-018-0306-6.

Received: 29 December 2017; Accepted: 6 November 2018; Published online: 7 January 2019

**References**

1. Castellano-Pozo, M. et al. R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Mol. Cell* 52, 583–590 (2013).

2. Nakama, M. et al. DNA–RNA hybrid formation mediates RNAi-directed heterochromatin formation. *Genes Cells* 17, 218–233 (2012).
3. Groh, M., Lufino, M. M., Wade-Martins, R. & Gromak, N. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet.* **10**, e1004318 (2014).

4. Sanz, I. A. et al. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol. Cell* **63**, 167–178 (2016).

5. Skourti-Stathaki, K., Kamieniarz-Gdula, K. & Proudfoot, N. J. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* **516**, 436–439 (2014).

6. Skourti-Stathaki, K., Proudfoot, N. J. & Gromak, N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* **42**, 794–805 (2011).

7. Powell, W. T. et al. R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. *Proc. Natl Acad. Sci. USA* **110**, 13938–13943 (2013).

8. Ginno, P. A. et al. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **45**, 814–825 (2012).

9. Arab, K. et al. Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A. *Mol. Cell* **55**, 604–614 (2014).

10. Barreto, G. et al. Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* **445**, 671–675 (2007).

11. Cortellino, S. et al. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67–79 (2011).

12. Schmitz, K. M. et al. TAF12 recruits Gadd45a and the nucleotide excision repair complex to the promoter of rRNA genes leading to active DNA demethylation. *Mol. Cell* **33**, 344–353 (2009).

13. Schäfer, A. et al. Ing1 functions in DNA demethylation by directing Gadd45a to H5K4me3. *Genes Dev.* **27**, 261–273 (2013).

14. Groh, M., Muecke, O. & Plass, C. R-loop-mediated genome instability is caused by impairment of replication fork progression. *Genes Dev.* **25**, 2041 (2011).

Acknowledgements
We acknowledge the support of O. Muecke and C. Plass in DNA methylation analysis, and the support of the DFKZ FACS and the IMB Microscopy and Genomics core facilities. We thank V. Vastolo for mouse embryos. I.G. was supported by the Helmholtz Foundation and by grants from the Deutsche Forschungsgemeinschaft (GR475/22-1, SFB1036), the CellNetworks Cluster of Excellence (EcTop 5), the DFKZ-MOST programme, and the Baden-Württemberg Stiftung (NCRNA_025). C.N. was supported by the European Research Council (ERC) DNA demethylase.

Author contributions
K.A. conceived the study, carried out most of the experiments and analysed the data. K.A., I.G., and C.N. designed the experimental work and wrote the manuscript. M.M. performed LC/MS-MS analyses, and E.K. performed the bioinformatics analyses. P.T. and A.S. generated and characterized HA-tagged Tet1 mESCs.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0306-6.
Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to K.A. or I.G. or C.N.
Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
Methods

Cell lines and culture conditions. Human PSFs, H387, HEK293T, HEK293TΔTet, and HEK293TΔTet cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM) containing 10% fetal calf serum, 5 mM L-glutamine, 100 U/ml penicillin and 100 μg/ml streptomycin. Mouse embryonic stem cells E14 (mESCs, E14tg2a) were grown on 0.1% gelatin-coated plates in DMEM, 20% FCS (Pan-Biotec), 0.4% leukemia inhibitory factor, 0.1 mM β-mercaptoethanol, 100 mM of MEM non-essential amino acids, 1 mM sodium pyruvate, 2 mM L-glutamine. Cell lines expressing N-terminal tagged FLAG-HA-TET1 in mouse embryonic cells were generated by electroporation of clamped short palindromic repeats (CRISPR)-CRISPR associated protein 9 (Cas9) technology. The single guide RNA (sgRNA) is designed using an online tool (see URL), cloned into pX330-U6-Chimeric, BB-CBI-hSpCas9 plasmid (Addgene number 42230) and co-transfected together with a repair template (sdDNA containing the FLAG-HA-sequence flanked by 60 bp Tetr (homology arms) and neomycin resistance plasmids into mESCs. G-418-resistant clones were genotyped and sequenced to select for homozygous integration. Expression of FLAG-HA-TET1 was confirmed by western blotting. Details of the sequences are listed in Supplementary Table 2. Overexpression of proteins or small interfering RNA (siRNA) knockdown in PSFs was carried out by electroporation (Amaxa). For other cell lines, transfections were carried out using TransIT (Mirus) for plasmids and RNAs. DharmaFect (Thermo Fisher Scientific) was used for transfection of siRNAs. Details on cell lines are listed in the Life Sciences Reporting Summary.

Antibodies, siRNAs and PCR primers. Antibodies used were: anti-FLAG (F7-425, Sigma), anti-FLAG (sc-7977, Santa Cruz), anti-HA (AB-9115, Santa Cruz), anti-TET1 (GTX-124207, Genetex), anti-DNMT3A (Ab-2851, Abcam) and anti-DNMT3B (NB-300-516, Novus Biologicals). The S9.6 antibody was purified from the hybridoma HB-8739 cell line. siRNAs against GADD45A, GADD45B, Rnase H1, Rnase H2a, Rnase H2b and Rnase H2c were purchased from Life Technologies, Ambion and Dharmacon. Plasmids used were: pcS2-FLAG, pGADD45A, pCDNA3-FLAG-PTB, pRTT-1 pG3DNA-3A-HA, pEGFP-N1-RnaseH1, pEGFP-N1-RnaseH1-B, pEGFP-N1-RnaseH1-H, pEGFP-N1-RnaseH1-H. Antibodies, siRNAs and PCR primers. Antibodies used were: anti-FLAG (F7-425, Sigma), anti-FLAG (sc-7977, Santa Cruz), anti-HA (AB-9115, Santa Cruz), anti-TET1 (GTX-124207, Genetex), anti-DNMT3A (Ab-2851, Abcam) and anti-DNMT3B (NB-300-516, Novus Biologicals). The S9.6 antibody was purified from the hybridoma HB-8739 cell line. siRNAs against GADD45A, GADD45B, Rnase H1, Rnase H2a, Rnase H2b and Rnase H2c were purchased from Life Technologies, Ambion and Dharmacon. Plasmids used were: pcS2-FLAG, pGADD45A, pCDNA3-FLAG-PTB, pRTT-1 pG3DNA-3A-HA, pEGFP-N1-RnaseH1, pEGFP-N1-RnaseH1-B, pEGFP-N1-RnaseH1-H, pEGFP-N1-RnaseH1-H. Antibodies used were: anti-FLAG (F7-425, Sigma), anti-FLAG (sc-7977, Santa Cruz), anti-HA (AB-9115, Santa Cruz), anti-TET1 (GTX-124207, Genetex), anti-DNMT3A (Ab-2851, Abcam) and anti-DNMT3B (NB-300-516, Novus Biologicals). The S9.6 antibody was purified from the hybridoma HB-8739 cell line. siRNAs against GADD45A, GADD45B, Rnase H1, Rnase H2a, Rnase H2b and Rnase H2c were purchased from Life Technologies, Ambion and Dharmacon. Plasmids used were: pcS2-FLAG, pGADD45A, pCDNA3-FLAG-PTB, pRTT-1 pG3DNA-3A-HA, pEGFP-N1-RnaseH1, pEGFP-N1-RnaseH1-B, pEGFP-N1-RnaseH1-H, pEGFP-N1-RnaseH1-H.

Cell cycle analysis. PSFs (5 x 10⁶) were synchronized by culturing for 17 h in the presence of 2 mM thymidine, for 9 h in fresh medium and for another 15 h in the presence of thymidine. Cells were transferred into fresh medium and collected at defined time points. Alternatively, cells were arrested with 2 mM thymidine for 25 h, released for 5 h by culturing in normal medium and treated for 8 h with 40 μg/ml nocodazole. For FACS sorting, 5 x 10⁶ cells were stained with 5 μM Hoechst 33342 (Sigma) for 1 h at 37°C and sorted using a BD-FACSaria II (BD Bioscience) cell sorter. We used 60,000 cells used for RNA and DNA analysis or for ChIP experiments. Details on flow cytometry can be found in the Life Sciences Reporting Summary.

ChIP assays. Cells were crosslinked with 1% formaldehyde for 10 min at room temperature and quenched with 0.5 M glycine for 5 min. After lysis and sonication to 250–800 bp DNA samples were transferred on G-418-resistant clones were genotyped and sequenced to select for homozygous integration. Expression of FLAG-HA-TET1 was confirmed by western blotting. Details of the sequences are listed in Supplementary Table 2. Overexpression of proteins or small interfering RNA (siRNA) knockdown in PSFs was carried out by electroporation (Amaxa). For other cell lines, transfections were carried out using TransIT (Mirus) for plasmids and RNAs. DharmaFect (Thermo Fisher Scientific) was used for transfection of siRNAs. Details on cell lines are listed in the Life Sciences Reporting Summary.

Antibodies, siRNAs and PCR primers. Antibodies used were: anti-FLAG (F7-425, Sigma), anti-FLAG (sc-7977, Santa Cruz), anti-HA (AB-9115, Santa Cruz), anti-TET1 (GTX-124207, Genetex), anti-DNMT3A (Ab-2851, Abcam) and anti-DNMT3B (NB-300-516, Novus Biologicals). The S9.6 antibody was purified from the hybridoma HB-8739 cell line. siRNAs against GADD45A, GADD45B, Rnase H1, Rnase H2a, Rnase H2b and Rnase H2c were purchased from Life Technologies, Ambion and Dharmacon. Plasmids used were: pcS2-FLAG, pGADD45A, pCDNA3-FLAG-PTB, pRTT-1 pG3DNA-3A-HA, pEGFP-N1-RnaseH1, pEGFP-N1-RnaseH1-B, pEGFP-N1-RnaseH1-H, pEGFP-N1-RnaseH1-H.

To monitor binding of GADD45A, PTB, GFP, TET1, and CBP to DNA-RNA hybrids, 2 x 10⁵ HEK293 cells overexpressing the tagged proteins were lysed in 100 μl 50 mM Tris-HCl, pH 7.6, 1 mM MgCl₂, 1 mM DTT, protein inhibitors Roche, 1 mM PMSF and 0.5% Triton X-100. After brief sonication, centrifugation and pre-clearing, 50 μl of lysate were incubated overnight with 10 μg 5’6’-amine group or mouse IgGs. Co-precipitated chromatin was left untreated or treated with RNase H1 (0.5 μg/ml) for 1 h at room temperature. After washing and elution, hybrid-associated proteins were monitored on immunoblots using anti-FLAG or anti-HA antibody.

ChIP–seq analysis of genomic TET1 occupancy. HA-TET1 tagged mESCs were permethylated in 15 ml cold PBS with 0.05% Tween 20 for 10 min at 4°C. After washing with PBS, cells were mock-treated or incubated with 0.2 μl RNase H1 (NEB) and 0.05 μg/ml 5’S1 nuclease (Fermentas) for 20 min at room temperature.

NATURE GENETICS | www.nature.com/naturegenetics
glycosylated with 4 U T4 phage beta-glycosyltransferase (T4-BGT) in 50 µl glycosylation buffer (30 mM potassium acetate, 20 mM Tris-acetate (pH 7.9), 10 mM magnesium acetate, 1 mM DTT, 100 mM uridine diphosphoraccharide (NEB) or left untreated. After incubation for 12 h at 37°C, T4-BGT was inactivated by incubation for 10 min at 75°C. DNA was digested withMsp1 (2 U, 6 h at 37°C), purified and analysed by qPCR. To normalize for digestion efficiency, the amount of amplicon corresponding to the Msp1 site (CpGr7) at the TCF21 promoter was normalized to the amplicon corresponding to the single Msp1 site on Cpg island 2. Relative levels of 5-hydromethylcytosine were calculated as a ratio of T4-BGT treated versus non-treated samples.

Base-resolution mapping of single-strand DNA footprint. Native bisulfite footprinting of R-loops was performed as described30, using the primers shown in Fig. 1c. In brief, 1 µg of DNA was treated with 3 M sodium bisulfite (pH 5.2) and 25 mM hydroquinone overnight at 37°C. After desulfonation with 0.3 M NaOH, bisulfite-modified DNA was purified and analysed by methylation-specific PCR using primers listed in Supplementary Table 2. Bisulfite-modified strand-specific primers together with unmethylated strand-specific primers were used for PCR. The PCR products were sequenced to determine the position and the size of the looped-out single-strand DNA.

In vitro transcription of TARID RNA. In vitro transcription assays consisted of 1 µg template DNA, 40 mM Tris-HCl (pH 7.9), 10 mM NaCl, 6 mM MgCl2, 10 mM DTT, 2 mM spermidine, 0.05% Tween 20, 0.5 mM each of ATP, GTP, CTP, UTP, 0.5 µg RNase inhibitor, and 200 U T7 RNA polymerase (Promega). After incubation for 16 h at 37°C and digestion with DNAse (5 U, 1 h at 37°C), the RNA was purified and quality-controlled by gel electrophoresis. Synthetic RNAs were transfected using TransIT (Mirus) reagent following manufacturer instructions.

Bioinformatic analysis. Sample demultiplexing and FastQ file generation were performed using illumina’s bcl2fastq conversion software v.2.18. The sequence reads were aligned to the mouse reference genome (NCBI37/mm9 build from illumina Genomes) using Bowtie v.2.2.8 (ref. 31) with options ‘--q -phred33 --end-to-end --maxins 1000’. Multi-mapped and discordant reads were removed with SAMtools v1.3.1 (ref. 32) with options ‘samtools view -f 2 -q 3’. Duplicated read pairs (likely PCR artefacts) were deduplicated using Picard MarkDuplicates v.2.6.0 (see URLs). The filtered BAM files were converted into RPKM-normalised bigwig coverage tracks using deepTools v.2.4.3 (ref. 33) and visualised together with previously published data using a custom UCSC browser session (see URLs). Metagene plots of average gene body read coverage were made with ngsplot1. Principal component analysis (PCA) of sample similarity was carried out with deepTools v.2.5.1. Peak calling for each ChIP sample against data sets were used for the identification of genomic R-loops4, TET1 (ref. 19) and TDG occupancy36. TARID-like lncRNAs were defined based on GENCODE M15 mouse gene annotation (see URLs) using the following two criteria: lncRNA exons located within 500 bp from the TSS of coding genes in antisense orientation; and detectable expression in mESCs from ENCODE (sample ENCSR000CWC) or from exosome mutants37. Feature coordinates were converted between the different mouse genome assemblies using the UCSC LiftOver utility (see URLs). HOMER v.4.9 transcription factor motif analysis38 was carried out using RNNH-sensitive TET1 peaks at TSS CGIs (n = 3,294) versus all other TET1 peaks at TSS CGIs (n = 7,588) as the matched background set. The same TET1 peak sets were used for plotting NRF1 heatmaps using deepTools v.2.5.1 and the processed ChIP-seq data of Domcke et al.12 (Gene Expression Omnibus (GEO) accession number GSE67867). Details on software and data deposition are listed in the Life Sciences Reporting Summary.

Statistical analysis. Data were tested for statistical significance by two-tailed t-tests using GraphPad Prism software and are shown as means of independent experiments (n) and standard deviation (± s.d.). significance level is indicated by asterisks (*P < 0.05, **P < 0.01, ***P < 0.001), or indicated as not significant (n.s.). Curve fitting for EMSA competition data points was generated using GraphPad Prism software. Enrichment analysis of TET1 peaks was performed in R using the one-sided Fisher’s exact test. A summary of statistical parameters and study design are listed in the Life Sciences Reporting Summary.

In vitro transcription of TARID RNA. In vitro transcription was carried out with deepTools v.2.5.1. Peak calling for each ChIP sample against data sets were used for the identification of genomic R-loops4, TET1 (ref. 19) and TDG occupancy36. TARID-like lncRNAs were defined based on GENCODE M15 mouse gene annotation (see URLs) using the following two criteria: lncRNA exons located within 500 bp from the TSS of coding genes in antisense orientation; and detectable expression in mESCs from ENCODE (sample ENCSR000CWC) or from exosome mutants37. Feature coordinates were converted between the different mouse genome assemblies using the UCSC LiftOver utility (see URLs). HOMER v.4.9 transcription factor motif analysis38 was carried out using RNNH-sensitive TET1 peaks at TSS CGIs (n = 3,294) versus all other TET1 peaks at TSS CGIs (n = 7,588) as the matched background set. The same TET1 peak sets were used for plotting NRF1 heatmaps using deepTools v.2.5.1 and the processed ChIP-seq data of Domcke et al.12 (Gene Expression Omnibus (GEO) accession number GSE67867). Details on software and data deposition are listed in the Life Sciences Reporting Summary.

In vitro transcription of TARID RNA. In vitro transcription was carried out with deepTools v.2.5.1. Peak calling for each ChIP sample against data sets were used for the identification of genomic R-loops4, TET1 (ref. 19) and TDG occupancy36. TARID-like lncRNAs were defined based on GENCODE M15 mouse gene annotation (see URLs) using the following two criteria: lncRNA exons located within 500 bp from the TSS of coding genes in antisense orientation; and detectable expression in mESCs from ENCODE (sample ENCSR000CWC) or from exosome mutants37. Feature coordinates were converted between the different mouse genome assemblies using the UCSC LiftOver utility (see URLs). HOMER v.4.9 transcription factor motif analysis38 was carried out using RNNH-sensitive TET1 peaks at TSS CGIs (n = 3,294) versus all other TET1 peaks at TSS CGIs (n = 7,588) as the matched background set. The same TET1 peak sets were used for plotting NRF1 heatmaps using deepTools v.2.5.1 and the processed ChIP-seq data of Domcke et al.12 (Gene Expression Omnibus (GEO) accession number GSE67867). Details on software and data deposition are listed in the Life Sciences Reporting Summary.

In vitro transcription of TARID RNA. In vitro transcription was carried out with deepTools v.2.5.1. Peak calling for each ChIP sample against data sets were used for the identification of genomic R-loops4, TET1 (ref. 19) and TDG occupancy36. TARID-like lncRNAs were defined based on GENCODE M15 mouse gene annotation (see URLs) using the following two criteria: lncRNA exons located within 500 bp from the TSS of coding genes in antisense orientation; and detectable expression in mESCs from ENCODE (sample ENCSR000CWC) or from exosome mutants37. Feature coordinates were converted between the different mouse genome assemblies using the UCSC LiftOver utility (see URLs). HOMER v.4.9 transcription factor motif analysis38 was carried out using RNNH-sensitive TET1 peaks at TSS CGIs (n = 3,294) versus all other TET1 peaks at TSS CGIs (n = 7,588) as the matched background set. The same TET1 peak sets were used for plotting NRF1 heatmaps using deepTools v.2.5.1 and the processed ChIP-seq data of Domcke et al.12 (Gene Expression Omnibus (GEO) accession number GSE67867). Details on software and data deposition are listed in the Life Sciences Reporting Summary.

In vitro transcription of TARID RNA. In vitro transcription was carried out with deepTools v.2.5.1. Peak calling for each ChIP sample against data sets were used for the identification of genomic R-loops4, TET1 (ref. 19) and TDG occupancy36. TARID-like lncRNAs were defined based on GENCODE M15 mouse gene annotation (see URLs) using the following two criteria: lncRNA exons located within 500 bp from the TSS of coding genes in antisense orientation; and detectable expression in mESCs from ENCODE (sample ENCSR000CWC) or from exosome mutants37. Feature coordinates were converted between the different mouse genome assemblies using the UCSC LiftOver utility (see URLs). HOMER v.4.9 transcription factor motif analysis38 was carried out using RNNH-sensitive TET1 peaks at TSS CGIs (n = 3,294) versus all other TET1 peaks at TSS CGIs (n = 7,588) as the matched background set. The same TET1 peak sets were used for plotting NRF1 heatmaps using deepTools v.2.5.1 and the processed ChIP-seq data of Domcke et al.12 (Gene Expression Omnibus (GEO) accession number GSE67867). Details on software and data deposition are listed in the Life Sciences Reporting Summary.

In vitro transcription of TARID RNA. In vitro transcription was carried out with deepTools v.2.5.1. Peak calling for each ChIP sample against data sets were used for the identification of genomic R-loops4, TET1 (ref. 19) and TDG occupancy36. TARID-like lncRNAs were defined based on GENCODE M15 mouse gene annotation (see URLs) using the following two criteria: lncRNA exons located within 500 bp from the TSS of coding genes in antisense orientation; and detectable expression in mESCs from ENCODE (sample ENCSR000CWC) or from exosome mutants37. Feature coordinates were converted between the different mouse genome assemblies using the UCSC LiftOver utility (see URLs). HOMER v.4.9 transcription factor motif analysis38 was carried out using RNNH-sensitive TET1 peaks at TSS CGIs (n = 3,294) versus all other TET1 peaks at TSS CGIs (n = 7,588) as the matched background set. The same TET1 peak sets were used for plotting NRF1 heatmaps using deepTools v.2.5.1 and the processed ChIP-seq data of Domcke et al.12 (Gene Expression Omnibus (GEO) accession number GSE67867). Details on software and data deposition are listed in the Life Sciences Reporting Summary.
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).

- 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 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

Our web collection on statistics for biologists may be useful.
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 | Methods |
|---------------------------------|---------|
| n/a                             | n/a     |
| ☑ Unique biological materials   | ☑ Involved in the study |
| ☑ Antibodies                    | ☐ Anti-HA rabbit polyclonal antibody (Reference: Ab-9110, company: Abcam). |
| ☑ Eukaryotic cell lines         | ☐ Anti-FLAG M2 mouse monoclonal antibody (Reference: F-7425, company: Sigma). |
| ☐ Palaeontology                 | ☐ Anti-TET1 (Reference: GTX-124207, company: Genetex). |
| ☑ Animals and other organisms   | ☐ Anti-DNMT3A (Reference: Ab-2850, company: Abcam). |
| ☑ Human research participants   | ☐ Anti-DNMT3B (Reference: NB-300-516, company: Novus Biologicals). |
| ☑ Human research participants   | ☐ Anti-GADD45A rabbit polyclonal (Reference: sc-797, company: Santa Cruz). |
| ☐ S9.6 antibody was purified from the hybridoma cell line HB-8739. | ☐ MRI-based neuroimaging |

Antibodies

| Antibodies used | Methods |
|-----------------|---------|
| 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 endogenously 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 authenticated 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 | 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 field-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 | 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_rep2.mm9_track.bw |
|                             | GSM2788889_Tet1_ChIP_Mock_rep2.mm9_peaks.bed.gz |
|                             | 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 | 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
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 (NCBI M37/mm9 build from Illumina iGenomes) using Bowtie v.2.2.8 with options ‘-q --phred33 --very-sensitive -p 4 --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 RPMK-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

- 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.