Published in final edited form as:
Nat Genet. 2019 February ; 51(2): 217–223. doi:10.1038/s41588-018-0306-6.

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

Khelifa Arab[1,2,3,*], **Emil Karaulanov**[1], **Michael Musheev**[1], **Philipp Trnka**[1], **Andrea Schäfer**[1], **Ingrid Grummt**[2,*], and **Christof Niehrs**[1,3,*]

[1]Institute of Molecular Biology (IMB), 55128 Mainz, Germany

[2]Division of Molecular Biology of the Cell II, German Cancer Research Center and DKFZ-ZMBH Alliance, 69120 Heidelberg

[3]Division of Molecular Embryology, German Cancer Research Center and DKFZ-ZMBH Alliance, 69120 Heidelberg

Abstract

R-loops are DNA:RNA hybrids enriched at CpG islands (CGIs), which can regulate chromatin states1–8. 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) directly binds to R-loops and mediates local DNA demethylation by recruiting TET1 (Ten-Eleven Translocation). Studying the tumor suppressor *TCF21*9, we find that antisense lncRNA TARID 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, which recruits the demethylation machinery to promoter CGIs.

We previously showed that the lncRNA TARID recruits TET enzymes to the tumor suppressor gene *TCF21*, leading to hypomethylation of the CpG island (CGI) promoter and increased *TCF21* expression9. TARID induces *TCF21* demethylation by recruiting

Contact: Christof Niehrs, Institute of Molecular Biology, Ackermannweg 4, 55128 Mainz, Germany Tel: +49(0)6131- 39-21400, c.niehrs@imb-mainz.de.
*joint corresponding authors

Data availability

TARID sequence information is deposited under GenBank: KF484512.1. The TET1 ChIP-Seq datasets are deposited in the Gene Expression Omnibus (GEO) repository (https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE104067.

Contributions

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

Competing financial interests

The authors declare no competing financial interests.

GADD45A, which targets TET and its cofactor Thymine DNA glycosylase (TDG) to specific genomic sites for DNA demethylation10–15.

TARID 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 *TARID,* which 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 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 HEK293^{TARIDwt} cells harboring an artificial *TCF21*-containing chromosome. No R-loops were observed in HEK293^{TARIDmut} 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* as well as *RPL13A* mRNA in PSFs and HEK293^{TARIDwt} cells. In contrast, overexpression of a catalytically inactive but binding-competent RNH1 mutant (HB) 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 footprinting16, which relies on the ability of sodium bisulfite to mutate cytosine (C) in single stranded DNA to uracil (C->U->T). Amplification of bisulfite-treated DNA using strand-specific primers matching the converted antisense DNA strand yielded products from TARID-expressing PSFs but not from TARID-deficient H387 cancer cells, indicating that the antisense strand is displaced by TARID (Fig. 1c). Consistently, sequencing of the PCR products from bisulfite-treated PSF DNA revealed C-to-T conversions in the region covering nucleotides –7 to +62, 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 TARID9. 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 (amplicon 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 analyzed 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/+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 demethylation9,10. We reasoned that GADD45A might directly interact with R-loops,

thereby targeting the demethylation machinery to the *TCF21* promoter. Indeed, ChIP-qPCR revealed that GADD45A occupancy was restricted to amplicon 2 comprising the R-loop (Fig. 2a). Overexpression of RNH1 reduced 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, indicating 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 retarded by GADD45A, while no binding to the labeled RNA oligonucleotide was observed (Fig. 2c and Supplementary Fig. 2d). Concordantly, in competitive EMSA assays, single- 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 (Figs. 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 via 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 TET114, suggesting that GADD45A may target TET1 to cellular R-loops. Indeed, both GADD45A and TET1 were co-immunoprecipitated with the S9.6 antibody, precipitation being abolished after RNH1 treatment (Fig. 2g and Supplementary Fig. 2f).

Simultaneous transcription of *TARID* and *TCF21* at the same time and from the same allele in both directions would lead to a head-to-head collision of RNA polymerases17. 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 FACS-sorted PSFs and HEK293^{TARIDwt} 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 G_2/M-phase cells (Fig. 3a and Supplementary Fig. 3a-b). Importantly, methylation of the *TCF21* promoter also oscillated, inversely correlating with TARID levels. In PSFs, the *TCF21* promoter was hypermethylated during G_1- and G_2/M-phase and hypomethylated in S-phase (Fig. 3b and Supplementary Fig. 3c). The promoter was almost fully demethylated in TARID-overexpressing HEK293^{TARIDwt} cells and in PSFs, while in *TARID*-deficient H387 and HEK293^{TARIDmut} 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 G_1/S boundary and released into the cell cycle for different times (Supplementary Fig. 3d). While TARID was transcribed in early/mid S-phase, *TCF21* mRNA was synthesized later, transcription starting in late S-phase (Fig. 3c). This sequential transcription of antisense and sense RNA was also observed in cells released from nocodazol-arrested G_2/M cells, corroborating 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 (Figure 3e). 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-hydroxymethylcyctosine (5hmC), the first oxidation intermediate of 5-methylcytosine (5mC), at the *TCF21* promoter (Fig. 3g). Concordantly, both TET1 occupancy and 5hmC levels at the *TCF21* promoter were elevated in S-phase (Fig. 3h). Conversely, RNH1 overexpression reduced 5hmC levels in S-phase (Fig. 3i). Binding of DNMT3A and -3B to the *TCF21* promoter was maximal in G_1-phase, inversely correlating with 5hmC levels (Supplementary Fig. 4d). These results suggest a sequential order of events, involving TARID-mediated R-loop formation, which guides GADD45A/TET1 to the *TCF21* promoter to activate transcription (Fig. 3j).

If R-loops are determinants for TET1-mediated DNA demethylation, changing the levels of R-loops should affect global levels of the oxidized 5-methylcytosine derivatives 5hmC, 5fC and 5caC. Indeed, knockdown of RNH1 or RNH2, which both degrade the RNA in R-loops18, increased TET1-dependent accumulation of demethylation intermediates (Fig. 4a and Supplementary Fig. 5a).

Next, we enriched R-loops from dissected mouse embryo heads and abdomen by immunoprecipitation with S9.6 antibody and analyzed 5mC oxidation derivatives. Strikingly, the levels of 5hmC, 5fC and 5caC were increased up to 10-fold in precipitated DNA, reinforcing that R-loops are sites of DNA demethylation *in vivo* (Fig. 4b).

We explored genomic TET1 occupancy following RNH1 treatment using ChIP-seq in mouse embryonic stem cells (mESCs), in which we homozygously tagged *Tet1* with hemagglutinin (HA). Tagged TET1 retained biological activity, that is, levels of oxidized 5-methylcytosine and expression of pluripotency markers were not affected (Supplementary Figs. 5b-d). TET1 ChIP-seq identified a total of 90,482 high confidence consensus peaks, which overlapped with the majority (67%) of previously published TET1 peaks 19 (Supplementary Figs. 6a-b). In RNH1-treated samples, 3,591 (~4%) of total TET1 peaks were decreased (FDR <0.01), indicating that a fraction of TET1 was associated with R-loops (Fig. 4c). Differential RNH1 sensitivity was confirmed by ChIP-qPCR for selected genes (Supplementary Fig. 6c).

Interestingly, 54% of RNH1-sensitive TET1 peaks overlapped with published genomic R-loops4 and showed strong preference (92%) for TSS-CGI regions (Figs. 4d-e and

Supplementary Table 1). The finding that TET1 binds to thousands of R-loop-containing TSS-CGIs in mESCs reinforces the view that these structures act as guides for TET demethylases. Among the RNH1-sensitive TET1 TSS-CGIs, 288 peaks associated with antisense lncRNAs (Supplementary Table 1, Supplementary Fig. 6d). As TARID is not conserved in mice, *TCF21* was not among these genes. However, *Gadd45a* itself was among the RNH1-sensitive TET1 peaks. Similar to the *TARID-TCF21* locus, the active *Gadd45a* promoter overlaps with an uncharacterized antisense lncRNA, suggesting an autoregulatory loop that modulates *Gadd45a* expression (Fig. 4f and Supplementary Fig. 6e). Motif analysis of *cis*-regulatory DNA elements revealed that RNH1-sensitive TET1 peaks are enriched in binding motifs for various transcription factors, the target sites of NRF1 being most prevalent (Supplementary Fig. 7a,b).

The presence of GADD45A was confirmed at 12 selected RNH1-sensitive TET1 peaks by ChIP-qPCR (Fig. 4g). Consistent with a fraction of TET1 being recruited to R-loop-containing promoters, GADD45A bound to RNH1-sensitive TET1 peaks but not to TET1 binding sites that were insensitive to RNH1 treatment (e.g. *Adamtsl5*, *Myocd, Ucp1*). DRIP-qPCR analysis confirmed that the position of R-loops is conserved at orthologous target genes in PSFs (Supplementary Fig. 8a). siRNA-mediated knockdown of GADD45A and GADD45B, which also promotes DNA demethylation20, decreased binding of TET1 to these target promoters (on average 40%), supporting that GADD45 promotes TET1 recruitment to R-loop-containing promoters but not to control promoters (Fig. 4h and Supplementary Fig. 8b).

In sum, our study shows that GADD45A is an epigenetic reader of promoter-associated R-loops. GADD45A may target also other factors to R-loops, such as TDG9,15. 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. While accumulation of R-loops is associated with genomic instability21,22, R-loops generated by lncRNAs may have a beneficial role in cell homeostasis, serving as sequence-specific barcodes for GADD45A/TET1 targeting.

Methods

Cell lines and culture conditions

Human primary skin fibroblasts cells (PSFs), H387, HEK293T, HEK293^{TARIDwt} and HEK293^{TARIDmut} cells9 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-Biotech), 0.4% leukemia inhibitory factor, 0.1 mM β-mercaptoethanol, 100 μM of essential amino-acids, 1mM sodium pyruvate, 5 mM L-glutamine. Cell lines expressing N-terminally tagged FLAG/HA-TET1 in mouse embryonic cells were generated with the CRISPR-Cas9 technology. The guide RNA (sgRNA) is designed using an online tool (see ^{URLs}), cloned into pX330-U6-Chimeric_BB-CBh-hSpCas9 plasmid (Addgene #42230) and co-transfected together with a repair template

Europe PMC Funders Author Manuscripts

(ssDNA containing the FLAG-HA-sequence flanked by 60 bp *Tet1* homology arms) and neomycine resistance plasmid 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 siRNAs knockdown in PSFs was carried 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 used are listed in the Life Sciences Reporting Summary.

Antibodies, siRNAs and PCR primers

Antibodies used were: anti-FLAG (F-7425, Sigma), anti-GADD45A (sc-797, Santa Cruz), anti-HA (Ab-9110, Abcam), 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 line23. siRNAs against, *GADD45A*, *GADD45B, RNase H1*, *H2a*, *H2b* and *H2c* were purchased from Life Technologies, Ambion and Dharmacon. Plasmids used were: pCS2^{+}-FLAG-GADD45A, pcDNA3-FLAG-PTB, FH-TET1-pEF, pcDNA3-HA-CBP, pEGFP-N1-RNaseH1, pEGFP-N1-RNaseH1-HB, pEGFP-N1-FLAG. DNA and RNA oligos were synthesized by Sigma are listed in Supplementary Table 2. Details on used antibodies are listed in the Life Sciences Reporting Summary.

RNA:DNA immunoprecipitation (DRIP)

DRIP assay was carried as described8. In brief, 20 μg DNA were treated with 10 ng/μl of RNase A for 5 min at RT, recovered by PCI extraction and digested with 4 U *EcoR*I, *EcoR*V, *Xba*I, *BamH*I and *Ssp*I (NEB) at 37°C for 3h or treated with RNase H1 (0.2 U/μl) for target validation of R-loops. 10 μg of fragmented DNA were incubated overnight with 10 μg S9.6 antibody in 0.1 ml IP buffer (20 mM HEPES-KOH [pH 7.5], 150 mM NaCl, 10 mM $MgCl_2$, 0.5% Triton X-100, 1mM dithiothreitol) and RNA:DNA hybrids were bound to protein A/G agarose (Sigma). After washing and treatment with Proteinase K (20 ng, 15 min at 45°C) DNA was purified and analyzed by qPCR.

To measure 5mC oxidation products in R-loops *in vivo*, abdomen and heads of 15.5 days mouse embryos were lysed in 500 μl lysis buffer (50 mM Tris-HCl [pH 7.5], 150 mM NaCl, 10 mM $MgCl_2$, 0.5% Triton X-100, 1mM dithiothreitol). Purified DNA was treated with RNase A (10 ng/100 μg total DNA) for 5 min at RT, purified and digested for 3h at 37°C with *EcoR*I, *EcoR*V, *Xba*I, *BamH*I and *Ssp*I. 30 μg of DNA were incubated overnight with 15 μg S9.6 antibody in 0.2 ml lysis buffer. RNA:DNA hybrids were recovered on protein A/G agarose and DNA was subjected to LC/MS-MS analysis.

URLs
Picard, http://broadinstitute.github.io/picard/
SeqMonk, http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/
UCSC genome browser and tools, http://genome-euro.ucsc.edu/
ENCODE blacklists, https://sites.google.com/site/anshulkundaje/projects/blacklists/
GENCODE annotation, https://www.gencodegenes.org/
Transcription factor motif enrichment analysis, http://homer.ucsd.edu/homer/motif/
Online CRISPR design tool, http://crispr.mit.edu

Cell cycle analysis

$5x10^6$ PSFs 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 ng/ml nocodazol. For FACS sorting, $5x10^6$ 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. 60,000 cells were used for RNA and DNA analysis or for ChIP experiments. Details on flow cytometry are listed in the Life Sciences Reporting Summary.

Chromatin immunoprecipitation (ChIP) assays

Cells were crosslinked with 1% formaldehyde for 10 min at RT and quenched with 0.5 M glycine. For GADD45A ChIP-qPCR candidate validation under RNase H1 pretreatment, the cells were permeabilized in 15 ml PBS containing 0.05% Tween 20 at 4°C for 10 min, washed with PBS, and incubated with 0.2 U/μl RNase H1 (NEB) and 0.05 U/μl S1 nuclease (Fermentas) for 20 min at RT. Samples were crosslinked with 0.75% formaldehyde for 15 min, quenched with 0.5 M glycine for 5 min, washed with PBS and crosslinked with 2 mM disuccinimidyl glutarate for 15 min. After washing, lysis and sonication, fragmented chromatin (200-600 bp) was diluted with IP buffer (20 mM Tris-HCl [pH 8.1], 200 mM NaCl, 1 mM EDTA, 1% Triton X-100), precleared for 1 h at 4°C with protein A/G agarose in the presence of 20 mg/ml sonicated salmon sperm DNA, and incubated overnight with the respective antibodies. Protein-DNA complexes were recovered on protein A/G Dynabeads (Thermo Fisher Scientific), followed by washes with lysis buffer, low salt buffer (200 mM NaCl, 50 mM Tris-HCl [pH 8.0], 5 mM $MgCl_2$, 1% Triton X-100), high salt buffer containing 500 mM NaCl, with LiCl buffer (250 mM LiCl, 100 mM Tris-HCl [pH 8.0], 5 mM EDTA, 0.5% Na-deoxycholate, 0.5% Triton X-100) and TE buffer. After reversal of the crosslink, DNA was purified, amplified by qPCR and normalized to input DNA and IgG-bound DNA.

To monitor binding of GADD45A, PTB, GFP, TET1 and CBP to DNA:RNA hybrids, $2x10^6$ HEK293 cells overexpressing the tagged proteins were lysed in 100 μl of 50 mM Tris-HCl [pH 7.6], 100 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, proteinase inhibitors (Roche), 1mM PMSF and 0.5% Triton X-100. After brief sonication, centrifugation and pre-clearing, 50 μl of lysate were incubated overnight with 10 μg S9.6 antibody or mouse IgGs. Co-precipitated chromatin was left untreated or treated with RNase H1 (0.5 U/μl) for 1 h at RT. 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-tagged mESCs were permeabilized 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 U/μl RNase H1 (NEB) and 0.05 U/μl S1 nuclease (Fermentas) for 20 min at RT. Cells were crosslinked with 2 mM disuccinimidyl glutarate (Sigma) for 45 min, re-crosslinked with 0.7% formaldehyde for 15 min at RT and quenched with 0.5 M glycine for 5 min. After lysis and sonication to 250-800 bp DNA fragmented chromatin was diluted with IP buffer (20 mM

Tris-HCl [pH 8.1], 200 mM NaCl, 1 mM EDTA, 1% Triton X-100), precleared for 1 h at 4°C on protein A/G agarose and incubated overnight with anti-HA antibody. Protein-DNA complexes were recovered on protein G Dynabeads (Thermo Fisher Scientific), followed by washes with low salt buffer (250 mM NaCl, 50 mM Tris-HCl [pH 8.0], 5 mM $MgCl_2$, 1% Triton X-100), with high salt buffer (500 mM NaCl, with LiCl buffer (500 mM LiCl, 100 mM Tris-HCl [pH 8.0], 5 mM EDTA, 0.5% Na-deoxycholate, 0.5% Triton X-100), and TE buffer. After reversal of the crosslink at 65°C overnight, DNA was treated with RNase A (2 ng/μl) for 1 h at 37°C and purified. Library preparation and amplification was performed according to the manufacturer's instructions (Ovation Ultralow V2, NuGEN) and sequenced on the NextSeq500 Illumina sequencer. ChIP-seq experiments were performed with three biological replicates employing one mESC clone (#4) harboring HA-tagged TET1 (see Supplementary Fig. 5b).

Pull-down assays, Northwestern blots and EMSAs

R-loops or DNA:RNA hybrid substrates were generated using either ^{32}P-end-labeled antisense RNA or sense DNA. For annealing, the RNA oligonucleotide and two complementary DNA oligonucleotides were heated for 5 min at 95°C in 50 μl annealing buffer (10 mM Tris–HCl, [pH 7.4], 1 mM EDTA). After slow cooling, probes were used in EMSA and pull-down assays. In pull-down assays, 10 μl beads containing 500 ng immobilized GADD45A were incubated for 1 h at RT with 2 pmoles of ^{32}P-labeled R-loop probes (2,000 cpm/ng) in 100 μl binding buffer (50 mM Tris-HCl [pH 7.9], 10% glycerol, 100 mM KCl, 5 mM $MgCl_2$, 10 mM β-mercaptoethanol, 0.1% NP-40). After washing and proteinase K digestion, released R-loops were analyzed by PAGE and PhosphorImaging.

For electrophoretic mobility shift assays (EMSAs) and Northwestern blots, FLAG-tagged proteins (GADD45A, PTB, GFP) were overexpressed in HEK293 cells and immobilized on FLAG epitope (M2) beads. After washing with 50 mM Tris-HCl [pH 7.9], 500 mM NaCl, 5 mM $MgCl_2$, 5 mM EDTA, 0.5% Triton X-100 and with 20 mM Tris-HCl [pH 7.9], 500 mM KCl, 5 mM $MgCl_2$, 0.5 mM DTT, 5 mM EDTA 1 M urea, beads were equilibrated in elution buffer (20 mM Tris-HCl [pH 7.9], 300 mM KCl, 5 mM $MgCl_2$, 0.5 mM DTT, 0.1 mM EDTA, 20% glycerol) and eluted in 100 μl elution buffer containing 0.05% NP-40 and 20 μg of 3X FLAG® peptide (Sigma) overnight at 4°C. Eluted Proteins were stored at -20°C.

For EMSAs, immunopurified GADD45A (5 pmoles/10 μl) was incubated for 1 h with 0.2 pmoles of labeled R-loop or hybrid probe (2,000 cpm/μl) in 20 mM Tris-HCl [pH 7.6], 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% glycerol, 100 μg/ml BSA. After electrophoresis on 0.8% agarose gels in 0.5xTAE buffer, samples were transferred to Hybond-N^+ membranes (GE Healthcare) and visualized by PhosphorImaging.

For Northwestern blots, FLAG-GADD45A, FLAG-PTB and FLAG-GFP were subjected to SDS-PAGE, transferred onto nitrocellulose membranes, renatured overnight at 4°C in binding buffer (10mM Tris-HCl [pH 7.5], 150 mM NaCl, 10 mM $MgCl_2$, 10 mM DTT, 0.5% Triton X-100) and incubated with 0.2 pmoles of the labeled R-loop probe (2,000 cpm/μl) for 3 h at RT. After washing and digestion with RNase H1 (50 U/ml) and RNase A (5 pg/ml) for 1 h, membranes were washed and subjected to PhosphorImaging.

Quantitative analysis of DNA methylation and hydroxymethyation

DNA methylation was analyzed on bisulfite-converted genomic DNA (EZ DNA Methylation Kit; Zymo Research) using the MassARRAY system (Sequenom) as reported24. Quantification of DNA methylation levels were calculated by the Epityper software v.1.2 (Sequenom). Primers are listed (Supplementary Table 2). For 5-hydroxymethylcytosine measurements, the assay was performed as reported9,25 with minor modifications. In brief, 250 ng genomic DNA was glycosylated with 4 U T4 phage beta-glucosyltransferase (T4-BGT) in 50 μl glycosylation buffer (50 mM potassium acetate, 20 mM Tris-acetate [pH 7.9], 10 mM magnesium acetate, 1 mM DTT, 100 mM uridine diphosphoglucose (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 with *Msp*I (2 U, 6 h at 37°C), purified and analyzed by qPCR. To normalize for digestion efficiency, the amount of amplicon corresponding to the *Msp*I site (CpG#7) at *TCF21* promoter was normalized to the amplicon corresponding to the single *Msp*I site on CpG island 2. Relative levels of 5-hydromethylcytosine were calculated as 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 described26, using the primers shown in Fig 1c. In brief, one microgram 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 analyzed by methylation-specific PCR using primers listed in Supplementary Table 2. Bisulfite-modified strand specific primers together with unmodified 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 contained 1 μg template DNA, 40 mM Tris-HCl [pH 7.9], 10 mM NaCl, 6 mM $MgCl_2$, 10 mM DTT, 2 mM spermidine, 0.05% Tween-20, 0.5 mM each of ATP, GTP, CTP, UTP, 0.5 U RNase inhibitor, and 40 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.

Liquid chromatography–mass spectrometry measurements of 5mC, 5hmC, 5fC, 5caC

LC-MS/MS analyses were performed as reported27. Briefly, 1 μg of genomic DNA or S9.6 immunoprecipitated DNA was digested with a mixture of nuclease P1 (Roche), snake venom phosphodiesterase (Worthington) and alkaline phosphatase (Fermentas). One hundred nanograms of digested DNA were subjected to LC-MS/MS analysis. Data were normalized using internal isotopic standards for each 5mC oxidation derivative.

Bioinformatic analysis

Sample demultiplexing and FastQ file generation was performed using Illumina's bcl2fastq conversion software v.2.18. The sequence reads were aligned to the mouse reference genome (NCBIM37/mm9 build from Illumina iGenomes) using Bowtie v.2.2.828 with options '-q --

phred33 --very-sensitive -p 4 --fr --end-to-end --maxins 1000'. Multi-mapped and discordant reads were removed with SAMtools v.1.3.129 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.2.4.330 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 ngs.plot31. Principal component analysis (PCA) of sample similarity was carried out with deepTools v.2.5.1. Peak calling for each ChIP sample against the respective pooled input control was performed with MACS v.2.1.132 and options 'macs2 callpeak -m 5 50 -g mm --bw 200 --format BAMPE'. Peaks overlapping abnormal genomic regions as determined by the ENCODE project (see URLs) were removed from further analysis using BEDTools v.2.25.033.

The filtered TET1 peaks from all six ChIP samples were imported in SeqMonk v.1.38.2 (see URLs). Using the tool "Contig Probe Generator" a consensus set of 90,482 peaks that were detected in two or more samples was defined and converted into a custom annotation track. Subsequently, the filtered BAM files of each ChIP sample were also loaded in SeqMonk using paired-end mode, thereby reconstituting the sequenced library fragments, and the raw fragment counts for each consensus TET1 peak were quantified. The generated count matrix was imported into R v.3.4.1 for differential enrichment analysis (RNH1- vs. mock-treatment) using the Bioconductor package DESeq2 v.1.16.1 with Independent Hypothesis Weighting (IHW) multiple testing procedure34,35. A conservative threshold of 1% false discovery rate (FDR) was applied to select robust differential peaks in RNH1-treated samples.

Subsets of TET1 peaks were analysed for overlap with published datasets using BEDTools and SeqMonk. Coordinates of CGIs and UCSC genes were downloaded from the UCSC mm9 Table Browser (see URLs). Published mouse ES cell datasets were used for the identification of genomic R-loops4, TET119 and TDG occupancy36. TARID-like lncRNAs were defined based on GENCODE M15 mouse gene annotation (see URLs) *employing the following two criteria: (i) lncRNA exons located within 500 bp from TSS of coding genes in antisense orientation and (ii) detectable expression in mouse ES cells 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 transcription factor motif analysis v.4.938 was carried out using RNH1-sensitive TET1 peaks at TSS-CGIs (n=3,294) vs. all other TET1 peaks at TSS-CGIs (n=7,588) as 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. 39 (GEO accession number GSE67867). Details on software and code as well as data deposition are listed in the Life Sciences Reporting Summary.

Statistical Analysis

Data were tested for statistical significance by two-tailed t-test using GraphPad Prism software and are shown as mean 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 were

generated using GraphPad Prism software. Enrichment analysis of TET1 peaks was performed in R using one-sided Fisher's Exact Test. A summary of statistical parameters and study design are listed in the Life Sciences Reporting Summary.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge the support of Oliver Muecke and Christoph Plass in DNA methylation analysis and the support of the DKFZ FACS- and the IMB Microscopy and Genomics core facilities. We thank Viviana 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 DKFZ-MOST program, and the Baden-Württemberg Stiftung (NCRNA_025). CN was supported by the ERC ("DNA Demethylase", "HybReader").

References

1. Castellano-Pozo M, et al. R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. Mol Cell. 2013; 52:583–590. [PubMed: 24211264]
2. Nakama M, et al. DNA-RNA hybrid formation mediates RNAi-directed heterochromatin formation. Genes Cells. 2012; 17:218–233. [PubMed: 22280061]
3. Groh M, Lufino MM, Wade-Martins R, Gromak N. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. PLoS Genet. 2014; 10
4. Sanz LA, et al. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. Mol Cell. 2016; 63:167–178. [PubMed: 27373332]
5. Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. R-loops induce repressive chromatin marks over mammalian gene terminators. Nature. 2014; 516:436–439. [PubMed: 25296254]
6. Skourti-Stathaki K, Proudfoot NJ, Gromak N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell. 2011; 42:794–805. [PubMed: 21700224]
7. Powell WT, et al. R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. Proc Natl Acad Sci USA. 2013; 110:13938–13943. [PubMed: 23918391]
8. Ginno PA, et al. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell. 2012; 45(6):814–825. [PubMed: 22387027]
9. Arab K, et al. Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A. Mol Cell. 2014; 55:604–614. [PubMed: 25087872]
10. Barreto G, et al. Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. Nature. 2007; 445:671–675. [PubMed: 17268471]
11. Cortellino S, et al. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell. 2011; 146:67–79. [PubMed: 21722948]
12. Schmitz KM, et al. TAF12 recruits Gadd45a and the nucleotide excision repair complex to the promoter of rRNA genes leading to active DNA demethylation. Mol Cell. 2009; 33:344–53. [PubMed: 19217408]
13. Schäfer A, et al. Ing1 functions in DNA demethylation by directing Gadd45a to H3K4me3. Genes Dev. 2013; 27:261–273. [PubMed: 23388825]
14. Kienhofer S, et al. GADD45a physically and functionally interacts with TET1. Differentiation. 2015; 90:59–68. [PubMed: 26546041]
15. Li Z, et al. Gadd45a promotes DNA demethylation through TDG. Nucl Acids Res. 2015; 43(8): 3986–3997. [PubMed: 25845601]

16. Sun Q, et al. R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science. 2013; 340(6132):619. [PubMed: 23641115]
17. Hobson DJ, Wei W, Steinmetz LM, Svejstrup JQ. RNA polymerase II collision interrupts convergent transcription. Mol Cell. 2012; 48:365–374. [PubMed: 23041286]
18. Cerritelli SM, Crouch RJ. Ribonuclease H: the enzymes in eukaryotes. FEBS. 2009; 276:1494–1505.
19. Wu H, et al. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature. 2011; 473:389–393. [PubMed: 21451524]
20. Ma DK. Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. Science. 2009; 323:1074–1077. [PubMed: 19119186]
21. Gan W, et al. R-loop-mediated genomic instability is caused by impairment of replication fork progression. Genes Dev. 2011; 25(19):2041. [PubMed: 21979917]
22. Bhatia V, et al. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. Nature. 2014; 511:362–365. [PubMed: 24896180]
23. Boguslawski SJ, et al. Characterization of monoclonal antibody to DNA.RNA and its application to imm*unodetection* of hybrids. J Immunol Methods. 1986; 89:123–130. [PubMed: 2422282]
24. Ehrich M, et al. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci USA. 2005; 102:15785–15890. [PubMed: 16243968]
25. Kinney SM, et al. Tissue-specific distribution and dynamic changes of 5-hydroxymethylcytosine in mammalian genomes. J Biol Chem. 2011; 286:24685–24693. [PubMed: 21610077]
26. Yu K, et al. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol. 2003; 4:442–451. [PubMed: 12679812]
27. Schomacher L, et al. Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. Nat Struct Mol Biol. 2016; 23:116–124. [PubMed: 26751644]
28. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9:357–359. [PubMed: 22388286]
29. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]
30. Ramirez F, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016; 44:W160–165. [PubMed: 27079975]
31. Shen L, Shao N, Liu X, Nestler E. Ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics. 2014; 15:284. [PubMed: 24735413]
32. Zhang Y, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008; 9:R137. [PubMed: 18798982]
33. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. [PubMed: 20110278]
34. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15:550. [PubMed: 25516281]
35. Ignatiadis N, Klaus B, Zaugg JB, Huber W. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. Nat Methods. 2016; 13:577–580. [PubMed: 27240256]
36. Neri F, et al. Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. Cell Rep. 2015; 10:674–683. [PubMed: 25660018]
37. Pefanis E, et al. RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. Cell. 2015; 161:774–789. [PubMed: 25957685]
38. Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010; 38:576–589. [PubMed: 20513432]
39. Domcke S, et al. Competition between DNA methylation and transcription factors determines binding of NRF1. Nature. 2015; 528:575–579. [PubMed: 26675734]

Europe PMC Funders Author Manuscripts

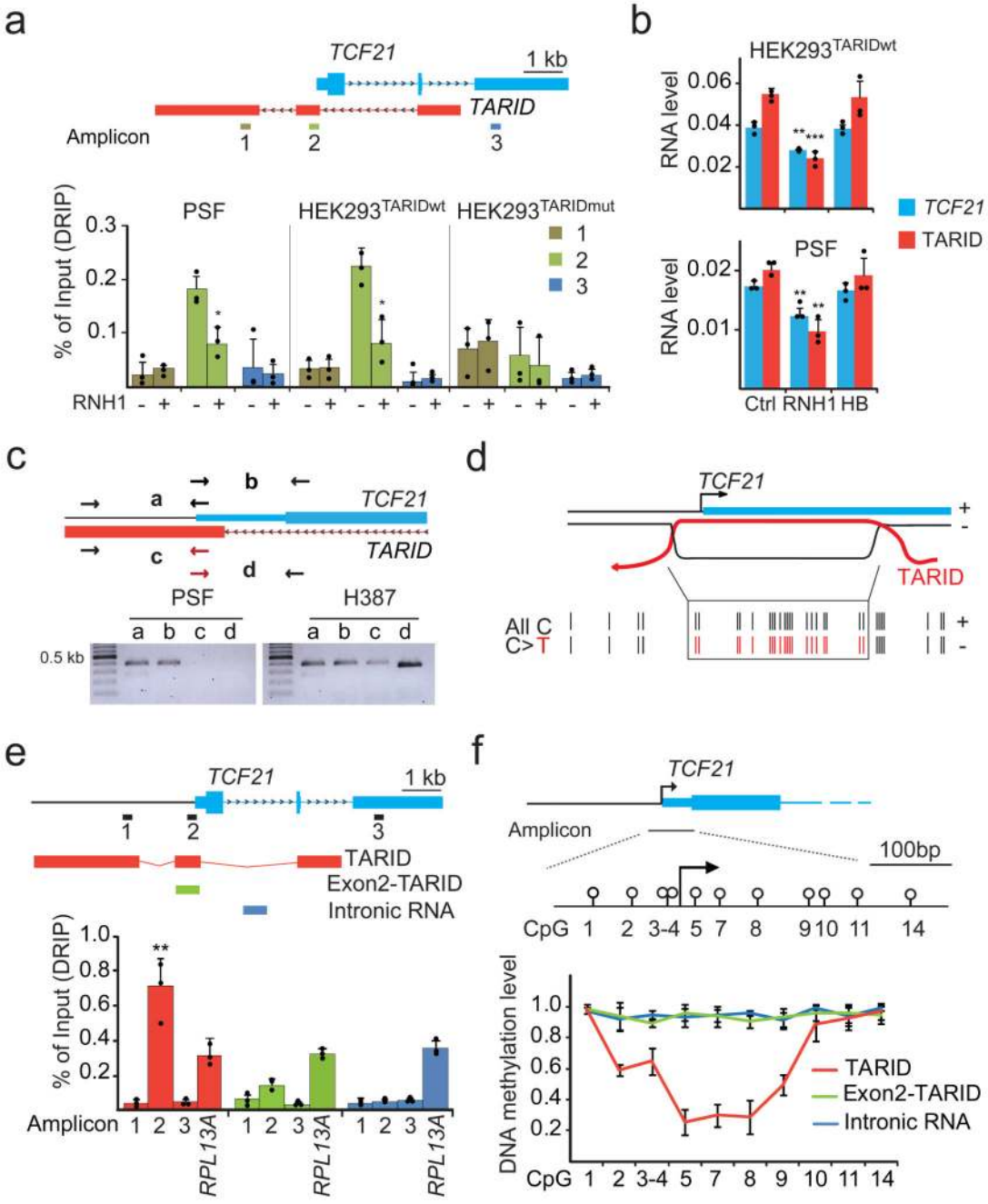

Figure 1. TARID forms an R-loop at the *TCF21* promoter.

a. DRIP-qPCR analysis at the *TCF21* promoter in the indicated cell lines transfected with plasmids encoding RNase H1 (RNH1 +) or control GFP (-). Scheme illustrates the *TCF21* locus and the position of amplicons used in qPCR.

b. RT-qPCR in cells transfected with plasmids encoding GFP (Ctrl), RNH1, or the hybrid-binding (HB) domain of RNH1. RNA levels were normalized to *HPRT1* mRNA.

Europe PMC Funders Author Manuscripts

c. R-loop footprinting showing PCR products of bisulfite-converted DNAs from PSF and H387 cells. Diagram depicts amplicons (a-d), position of unmodified primers (black), and primers matching bisulfite-modified DNA (red) used for PCR.

d. Native bisulfite sequencing of single-stranded DNA showing TARID-R-loop position at the *TCF21* transcription starting site (see also Supplementary Fig. 1d). Cartoon indicating in scale the position of cytosines (all C) and of Cs converted to T residues (R-loop position) in the region analyzed.

e. Ectopic TARID forms an R-loop at *TCF21*. DRIP-qPCR in H387 cells transfected with TARID full-length (red*),* exon2 (green*),* or intronic control RNA (blue). *RPL13A*, R-loop positive control.

f. TARID induces demethylation of the *TCF21* promoter. MassARRAY DNA methylation analysis in H387 cells transfected with TARID variants. The position of CpG residues is indicated by open lollipops. Methylation level is scaled from 0 to 1 and plotted for each CpG residue in the amplicon.

Panels **a**,**b**,**e**,**f**: mean ± s.d., n = 3 biological replicates, two tailed t-test; **P*<0.05, ***P*<0.01, *** *P*<0.001. Panels **c**,**d**: Experiments were repeated once with similar results. Blot image was cropped (see Supplementary Fig. 9a).

Europe PMC Funders Author Manuscripts

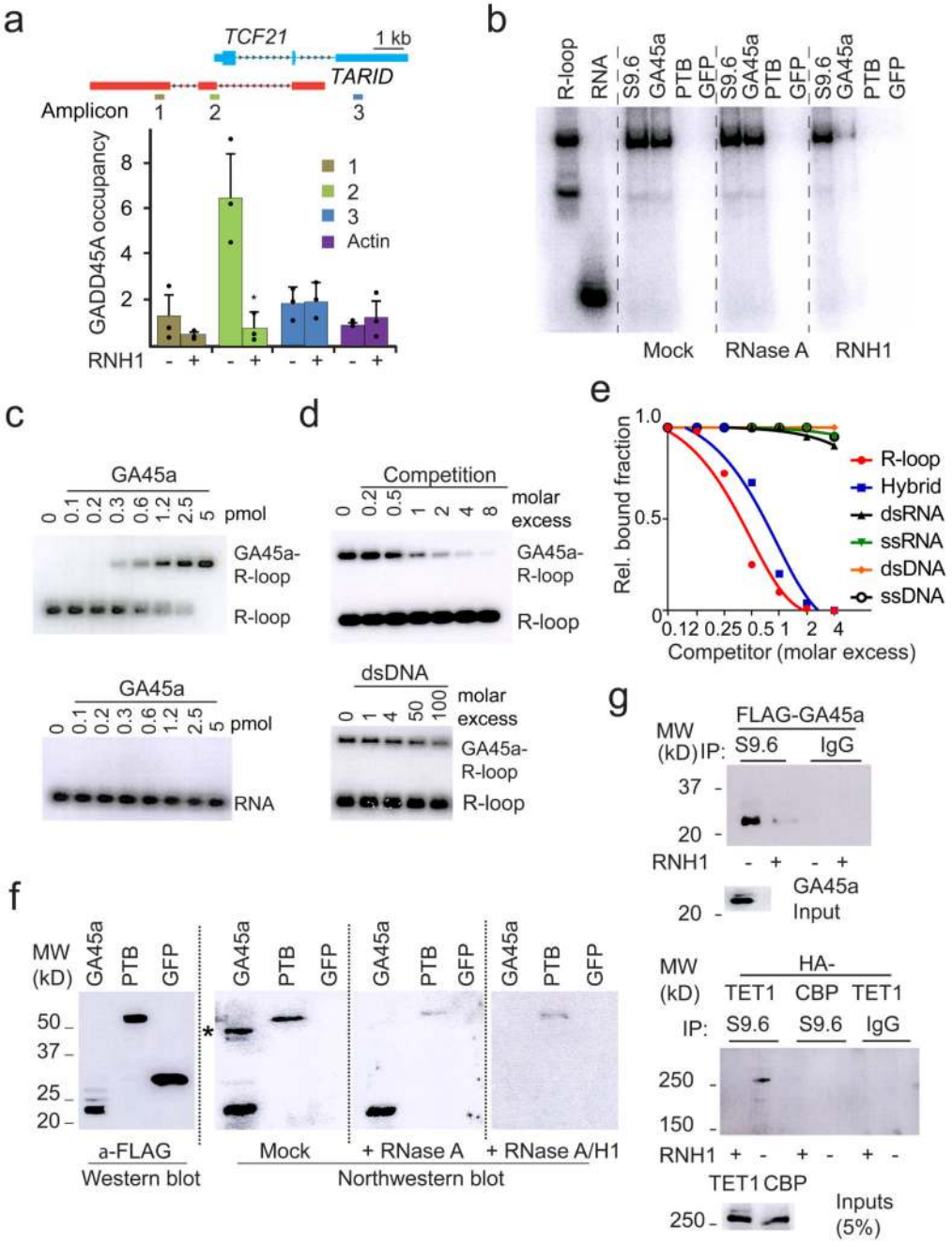

Figure 2. GADD45A binds R-loop structures *in vitro* and *in vivo*.

a. ChIP-qPCR of GADD45A at the *TCF21* promoter in PSFs overexpressing RNase H1 (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.

Europe PMC Funders Author Manuscripts

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 dsDNA (lower panel).

e. Quantitation of competitive EMSAs using a GADD45A/R-loop complex and the indicated unlabeled competitors. GADD45A binding to R-loops was set to 1.

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 RNase H1 (RNH1) before immunoblotting.

Panels **b**-**g**: Experiments were repeated twice with similar results. All blot images were cropped (see Supplementary Figs. 9b-d and 10a-b).

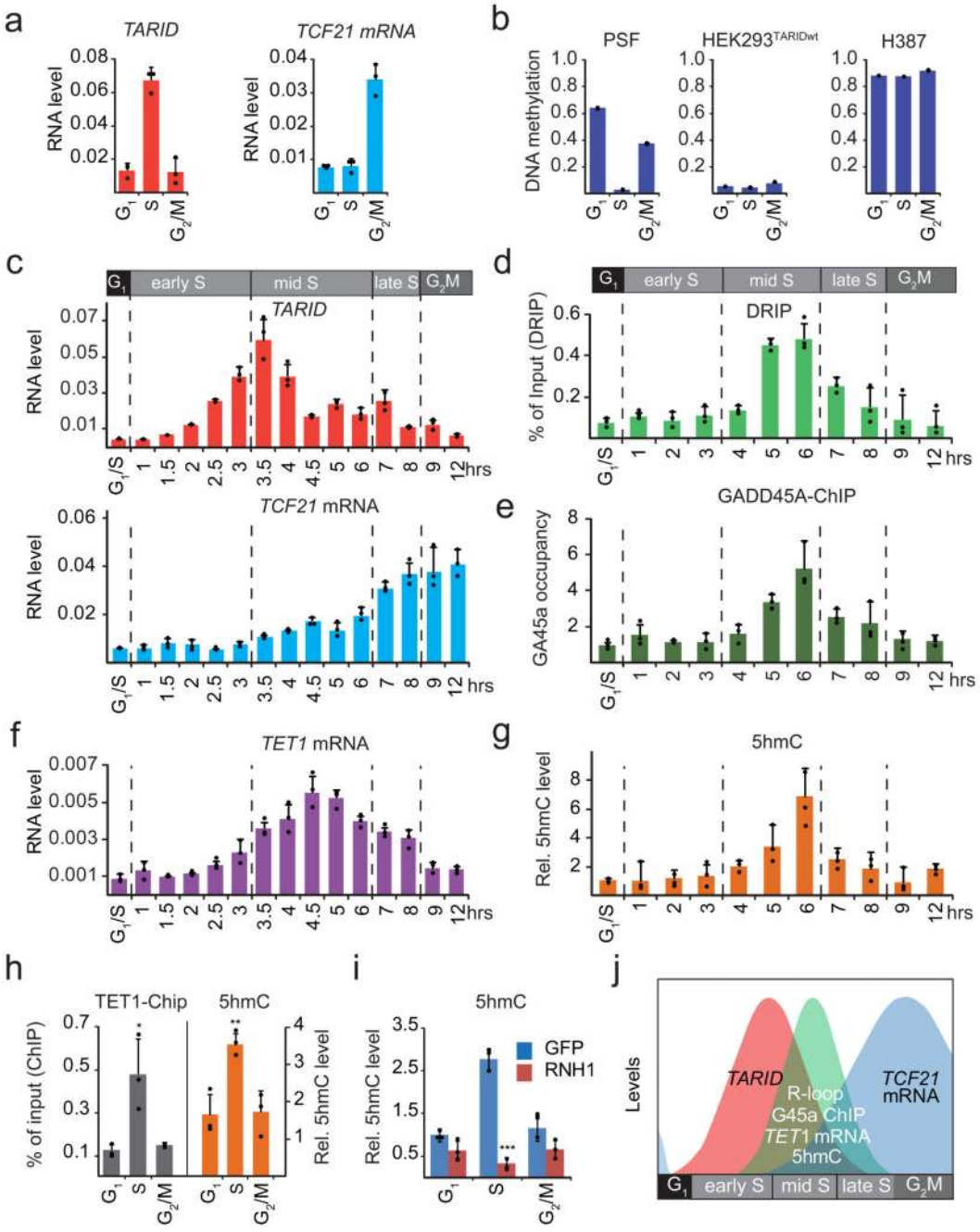

Figure 3. Transcription of *TARID* precedes R-loop formation and *TCF21* activation.

a. RT-PCR analysis of TARID and *TCF21* mRNA in FACS-sorted PSFs. RNA levels were normalized to *HPRT1* mRNA.

b. DNA methylation at *the T*SS of *TCF21* measured *by* MassARRAY *i*n FACS-sorted cells. Methylation is plotted as mean value of all informative CpG residues in the amplicon scaled from 0 to 1.

Europe PMC Funders Author Manuscripts

c. TARID (red) and *TCF21* mRNA (blue) levels in PSFs arrested at the G_1/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 GADD45A occupancy at the *TCF21* promoter in synchronized PSFs. The levels are normalized to control IgGs.

f. RT-qPCR monitoring *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 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 12h following transfection with RNaseH1 (RNH1) or GFP.

j. Schematic of peaks of TARID expression, R-loop formation, and *TCF21* expression proceeding sequentially during cell cycle progression.

Panels **a**-**i**: mean ± s.d., n = 3 biological replicates; **h**-**i** two tailed t-test; **P*<0.05, ***P*<0.01, *** *P*<0.001.

Europe PMC Funders Author Manuscripts

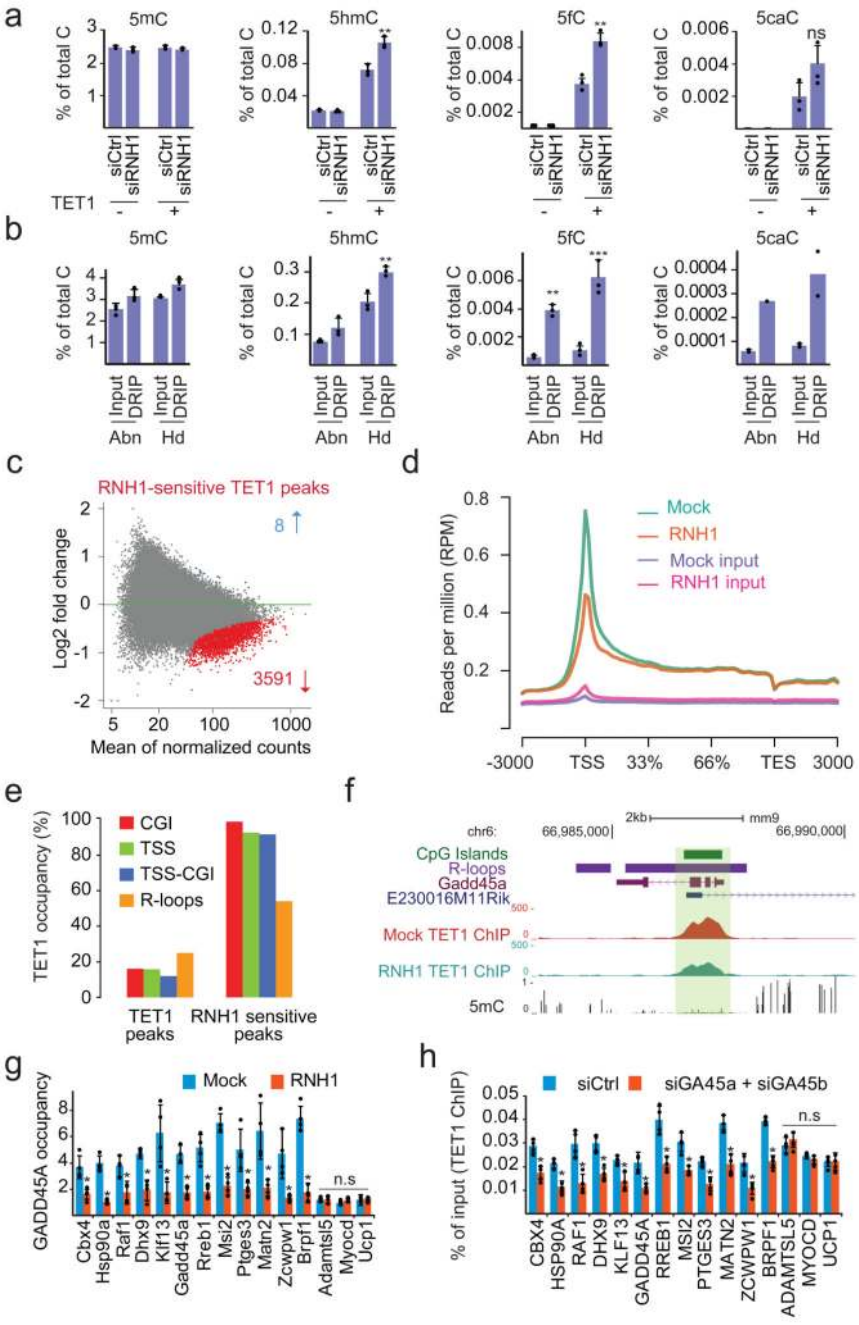

Figure 4. R-loops recruit TET1 to CGI promoters.

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 stage E15.5.

Europe PMC Funders Author Manuscripts

c. Differential enrichment TET1 ChIP-seq analysis of RNH1- vs. mock-treated mESCs 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 vs. RNH1-sensitive TET1 peaks with CpG islands (CGI), Transcription Start Sites (TSS), and published R-loops4.
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-loops4, and CpG methylation. A broader view of this locus with all ChIP-seq samples is shown in Supplementary Fig. 6e.
g. ChIP-qPCR monitoring GADD45A occupancy at RNase H1-sensitive TET1 peaks in the absence (Mock) or presence of RNase H1 (RNH1) along with negative controls. Data are normalized to control IgGs.
h. ChIP-qPCR monitoring TET1 binding in human primary skin fibroblasts at promoters of human orthologs of the genes shown in (**g**) upon treatment with GADD45A- and GADD45B-specific siRNA or scrambled control siRNAs.
Panels **a**,**b**,**g**,**h** : mean ± s.d., n = 3 (**g**,**h** n=4;) biological replicates, two tailed t-test; **P*<0.05, ***P*<0.01, *** *P*<0.001.