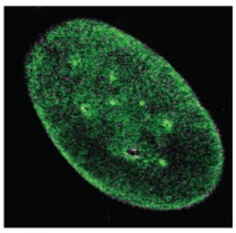
Nucleus



Nucleus

ISSN: 1949-1034 (Print) 1949-1042 (Online) Journal homepage: www.tandfonline.com/journals/kncl20

TET dioxygenases localize at splicing speckles and promote RNA splicing

Florian D. Hastert, Jasmin Weber , Christina Bauer , Andreas Zhadan , Deepanshu N. D. Singh , Thomas C. Dix , Roland Arnold , Sergey Bessonov , Matthias Soller , Heinrich Leonhardt , M. Cristina Cardoso & Maria Arroyo

To cite this article: Florian D. Hastert, Jasmin Weber , Christina Bauer , Andreas Zhadan , Deepanshu N. D. Singh , Thomas C. Dix , Roland Arnold , Sergey Bessonov , Matthias Soller , Heinrich Leonhardt , M. Cristina Cardoso & Maria Arroyo (2025) TET dioxygenases localize at splicing speckles and promote RNA splicing, Nucleus, 16:1, 2536902, DOI: 10.1080/19491034.2025.2536902

To link to this article: https://doi.org/10.1080/19491034.2025.2536902

NUCLEUS
2025, VOL. 16, NO. 1, 2536902
https://doi.org/10.1080/19491034.2025.2536902

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

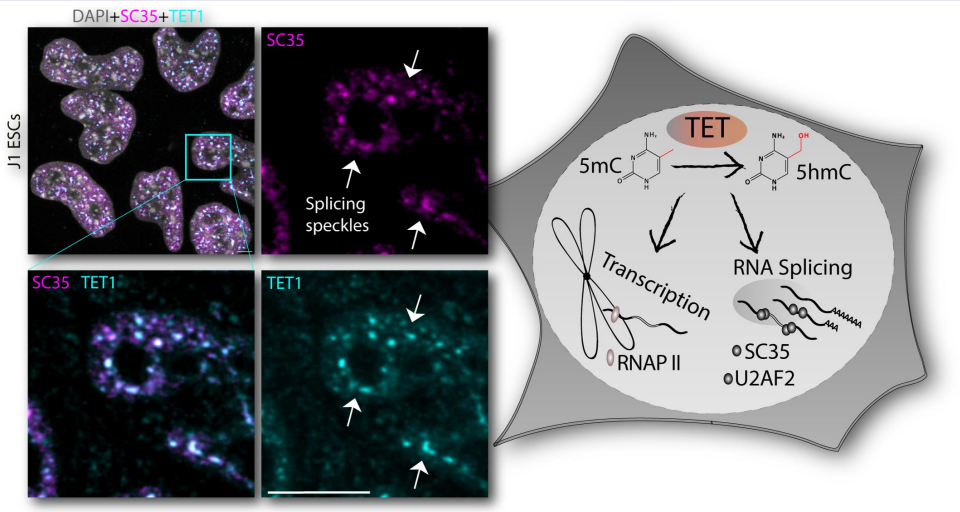
TET dioxygenases localize at splicing speckles and promote RNA splicing

Florian D. Hastert [a*], Jasmin Weber [a#], Christina Bauer [b^], Andreas Zhadan [a], Deepanshu N. D. Singh [c], Thomas C. Dix [c], Roland Arnold [c,d], Sergey Bessonov [e*], Matthias Soller [c,f], Heinrich Leonhardt [b], M. Cristina Cardoso [a], and Maria Arroyo [a]

[a]Cell Biology and Epigenetics, Department of Biology, Technical University of Darmstadt, Darmstadt, Germany; [b]Faculty of Biology and Center for Molecular Biosystems (BioSysM), Human Biology and BioImaging, LMU Munich, Munich, Germany; [c]School of Biosciences, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham, UK; [d]Department of Cancer and Genomics Sciences, School of Medical Sciences, College of Medicine and Health, University of Birmingham, Edgbaston, Birmingham, UK; [e]Department of Cellular Biochemistry, Max Planck Institute for Multidisciplinary Sciences, Göttingen,, Germany; [f]Division of Molecular and Cellular Function, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

ABSTRACT

The dynamic regulation of RNA metabolism plays a crucial part in cellular function, with emerging evidence suggesting an important role for RNA modifications in this process. This study explores the relationship between RNA splicing and the TET dioxygenase activity, shedding light on the role of hm5C (RNA 5-hydroxymethylcytosine), and TET proteins in RNA metabolism. Integrating data from mass spectrometry, AlphaFold structural modeling, microscopic analysis, and different functional assays, including in vitro splicing, TET proteins were found to regulate splicing. We show that TET1, TET2, and TET3 interact with the splicing factors U2AF1 and U2AF2. Interestingly, TET dioxygenases localize in splicing speckles in mammalian and Drosophila cells. TET speckles association was found to be RNA-dependent, but also rely on its interaction with splicing factors. Furthermore, cellular splicing assays revealed that all three TET proteins promote splicing efficiency independent of their catalytic activity. Interestingly, though, the oxidation of m5C to hm5C restores splicing efficiency in vitro. The latter highlights the regulatory role of cytosine modifications in RNA metabolism. These findings provide insights into the complex interplay between RNA modifications and splicing, suggesting a multifaceted role for TET proteins in RNA metabolism beyond their canonical function in the oxidation of 5mC in DNA.



ARTICLE HISTORY
Received 1 April 2025
Revised 15 July 2025
Accepted 16 July 2025

KEYWORDS
5-hydroxymethylcytosine; epitranscriptomics; RNA modifications; splicing; splicing speckles; TET dioxygenases; U2AF

CONTACT M. Cristina Cardoso cardoso@bio.tu-darmstadt.de Cell Biology and Epigenetics, Department of Biology,Technical University of Darmstadt Institution, Darmstadt 64287, Germany; Maria Arroyo arroyo.lopez.mc@gmail.com Cell Biology and Epigenetics, Department of Biology, Technical University of Darmstadt, Darmstadt 64287, Germany
*Present address: BioNTech SE 55,131 Mainz, Germany.
#Present address: LETI Pharma GmbH, Schnittspahnstraße 10, 085,737 Ismaning, Germany.
^Present address: BÜHLMANN Laboratories AG,Schönenbuch, Switzerland.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/19491034.2025.2536902

Introduction

Cellular ribonucleic acids (RNAs) display many biological functions, and the RNA processing machinery is crucial for the normal function of a cell. Protein-coding RNAs need to be processed to remove non-coding regions, a task performed by a macromolecular machine, the spliceosome. One of its components is SC35, belonging to the family of serine-arginine-rich proteins (SR proteins), which is involved in intron recognition and spliceosome assembly [1]. Its localization is mainly restricted to so-called nuclear speckles, also referred to as 'SC35 domains' [2], RNA – protein interchromatin granules found in the nucleoplasm of eukaryotic cells. The term 'speckles' was first used in 1961, and 2 years later, these speckles were identified as RNA-containing interchromatin particles [3]. Speckles are enriched for RNA-binding proteins (RBPs) that belong mainly to the different subsets of pre-mRNA splicing factors [4–6]. Interestingly, SC35 distribution is unaffected by RNase A digestion, indicating that the association with speckles is based on protein–protein interactions rather than protein–RNA interactions [4]. Besides the SR protein family, various other splicing factors contain SR domains. One of them is the U2 small ribonucleoprotein particle (snRNP) auxiliary factor 2 (U2AF2), with an N-terminal SR domain and three C-terminal RNA-binding domains [7,8]. Its structure is conserved throughout different species, including *Drosophila melanogaster* (dU2AF50) and *Saccharomyces cerevisiae* (Mud2) [9]. U2AF2 interacts with a smaller subunit called U2AF1, also containing an SR domain, which was found to interact with SC35 [10]. During splicing, the freshly transcribed premature messenger RNA (pre-mRNA) undergoes a series of rearrangements, which can occur even during ongoing transcription. This is a fundamental process that increases the diversity of the transcriptome and, consequently, the proteome. The spliceosome complex plays a major role in its regulation, and its importance is underlined by aberrant splicing events that were associated with different diseases, including diabetes, stroke, hypertension, and cancer [11–13]. Important roles for splicing are also related to biological processes like terminal erythropoiesis [14] or neurogenesis [15]. Furthermore, alternative splicing events increase transcript diversity in different cell types and under variant conditions [15]. Internal base editing and splicing guarantee the exclusive continuance of coding sections, while distinct modifications at the 5′ and 3′ untranslated regions (UTR) ensure mRNA stability and correct subcellular localization [16]. Translational capacity is also dependent on RNA modifications like the incorporation of pseudouridine [17] or N6-methyladenosine (m6A) [18]. Recently, it has been shown that splicing of mRNA is more efficient when tethered to the transcription elongation complex, as the newly synthesized RNA strand is extruded from RNA polymerase II (RNAPII) [19].

In mammals, it is well established that DNA cytosine base modifications, especially 5mC, are important epigenetic marks and fulfill crucial functions in cellular processes [20]. 5mC is generated and maintained by DNA methyltransferases (MTases) called Dnmts [21] and is mainly found in CpG dinucleotide-rich sites, resulting in dense packing of chromatin [22]. In 1978, m5C was also detected in the rRNA of wheat seedlings [23], raising the question of whether it can fulfill regulatory functions in RNA. Since then, it has been shown that m5C contributes to the structural stabilization of RNA molecules, impairing the recognition of endogenous RNAs by the innate immune system [24]. In lncRNAs, m5C can influence their binding dynamics to chromatin-modifying complexes. Furthermore, it can take part in the generation of individual sets of mRNAs by affecting the metabolism of microRNAs [25]. In tRNA and rRNA, m5C has been extensively studied, but little is known about its role in mRNA, in part, due to the lack of effective separation and purification technologies [26]. Similar to DNA, methylation of cytosines in RNA is catalyzed by specific RNA MTases (RMTases) like TRDMT1 and the nucleolar protein 1 (NOL1)/NSUN protein family [27,28]. These proteins use S-adenosylmethionine (SAM) as a methyl group donor [24]. However, the identity of m5C

modifiers and erasers is still vague. Since the discovery of m5C, more than 300 RNA modifications have been identified [29,30], N6-methyladenosine (m6A) being one of the most relevant. This modification was shown to be the substrate for iterative oxidation to N6–hydroxymethyl adenosine and further to N6-formyladenosine, in a reaction catalyzed by the AlkB protein family of dioxygenases [31,32]. Notably, members of the AlkB family and the DNA modifying ten-eleven-translocation (TET) family share structural similarities. TET proteins are Fe(II) and 2-oxoglutarate-dependent dioxygenases and are known to act on DNA by oxidizing 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) and further to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). The mammalian TET family consists of three members, TET1, TET2, and TET3, sharing a conserved catalytic domain at the C-terminus [33]. Interestingly, TET1 is the most ubiquitously expressed member across different tissues [34]. The AlkB family of dioxygenases can use DNA and RNA as substrates [35,36]; however, whether the related TET proteins do the same was unclear. Later on, it was shown that RNA can be the substrate of TET dioxygenases, as the occurrence of RNA 5-hydroxymethylcytosine, termed hm5C, is dependent on the presence of TET proteins [37]. Furthermore, hm5C has been detected in different tissues, among others, the brain, heart, pancreas, and spleen. The iterative oxidation of m5C to hm5C in total RNA was demonstrated in a mouse model, suggesting a conserved process that could have critical regulatory functions inside cells [38]. In addition, 5-formylcytidine (f5C) was detected in mammalian RNA, where its formation is currently assumed to be catalyzed by TET proteins [39]. Interestingly, TET has been shown to interact with microRNAs regulating hm5C in the adult brain [40], and hm5C was found in the RNA of Drosophila S2 cells [41]. These cells contain a conserved TET analog (called dTET), which favors the conversion of m5C to hm5C in polyadenylated RNA, resulting in a more efficient translation of the hydroxymethylated mRNA.

Although the function of TET proteins in the oxidation of m5C seems clear, a lot about the role of RNA modifications and their effect on RNA metabolism remains elusive. Splicing has emerged as a key regulator of neural development, where spliceosome dysfunction causes a series of neurodevelopmental disorders with similar features [42]. In this context, investigating the so-called 'splicing code' and novel factors involved in splicing regulation has gained interest. In the present study, we address the connection between RNA splicing and TET dioxygenases and their effect on RNA splicing. We show that TET proteins localize to splicing speckles in mouse, human, and Drosophila cells and tissues. This association with speckles is RNA-dependent, as is their interaction with splicing factors underlined. Furthermore, we show that m5C blocks splicing, while its oxidative derivative, hm5C, restores splicing activity. Overall, our results provide novel insights into how RNA splicing is regulated, with TET proteins and their oxidation product hm5C having an impact on it.

Materials and methods

Expression constructs

Mammalian expression plasmid encoding mCherry-SC35 (pc3332) was generated by cutting the SC35 coding sequence (CDS) from the vector pEYFP-SC35 (pc1202) and cloned into the pmCherry-C1 backbone (pc2387) using restriction enzymes BspE1 and HindIII. The vectors pWHE237 (pc3336) and pWHE237mod-WT (pc3337) were used for luciferase reporter gene assays (generated by C. Berens´s laboratory) and published in [43]. Both vectors carry the luciferase (luc+) gene. The intron 2 of the human β-globin gene (bgl2 intron) was inserted into the luc+ gene (30 nt downstream of the start codon using the sequence 5'-GTGAGT//CCACAG-3' as an intron boundary). To generate the splicing reporter vector phnRNP-DL-EGFP_ECFP (pc3351), the firefly luciferase sequence was cut from the vector pWHE200 (pc3335) using BamHI and NotI restriction enzymes. The ECFP reverse complement CDS was amplified from pc0911 using following primers: forward 5'-gagtactgcagttacttgtacagc-3," reverse 5"-

gagtagaattcatggtgagcaagg-3.' The Renilla luciferase sequence was then replaced using PstI and EcoRI restriction. All other expression constructs have been described and published in previous studies and can be found in Supplementary Table S1.

Cell culture, embryoid body formation, and transfection

Cells were cultured in DMEM (Cat.No.: 41965039, Gibco, Massachusetts, USA) supplemented with fetal bovine serum (FBS) 10% (Capricorn Scientific, Cat.No.: FBS-22A), 2 mM L-glutamate (VWR, Cat.No.: 56–85–9), 0.1% gentamicin (Sigma-Aldrich, Cat.No.: 1405–41–0), and sodium pyruvate 110 mg/L (Sigma-Aldrich, Cat.No.: 113–24–6) in an cell culture incubator at 37°C and 5% CO_2. Mouse embryonic fibroblasts (MEF), provided by Jenuwein Laboratory (Freiburg, Germany) [44], were cultured in DMEM containing 15% FBS [45]. J1 wild-type mESCs were derived from a male agouti 129/terSv embryo generated by [46], provided by Leonhardt Laboratory (LMU, Munich, Germany), together with J1 Tet mutants, and Dnmt TKO cell lines. ES v6.5 wt and Tet TKO cell lines [47] were also provided by Leonhardt Laboratory. The ESC medium for J1 mESCs was supplemented with 16% fetal bovine serum (FBS) (Capricorn Scientific, Cat.No.: FBS-22A), 1x nonessential amino acids (Cat. No.: M7145, Sigma-Aldrich, St Louis, MO, USA), 1x penicillin/streptomycin instead of gentamicin (Cat. No.: P4333, Sigma-Aldrich, St Louis, MO, USA), 2 mM L-glutamine (VWR, Cat.No.: 56–85–9), 0.1 mM β-mercaptoethanol (Cat. No.: 4227, Carl Roth, Karlsruhe, Germany), 1000 U/ml LIF (Cat. No.: ESG1107, Merck, Kenilworth, NJ, USA), 1 µM of MEK inhibitor PD0325901 and 3 µM of GSK3 inhibitor CHIR99021 (2 inhibitors (2i), Axon Medchem BV, Reston, VA, USA). Stem cells were maintained in a naïve pluripotent ground state by culturing them under feeder-free conditions on gelatin-coated culture dishes (0.2% gelatin in ddH_2O, Sigma-Aldrich Chemie GmbH, Steinheim, Germany, Cat. No.: G2500) in the above-described stem cell medium. All cells were frozen in freezing medium (DMEM supplemented with 20% FBS, 50 µg/ml gentamicin, and 2 mM L-glutamate, 10% DMSO), and regularly tested for mycoplasma to ensure they were contamination-free. Depending on the experiment, cells were split every two to three days except for ESCs, which were split every 2 days on gelatinized culture dishes. For in vitro embryoid body (EB) formation, J1 ESCs were trypsinized and approximately 37,500 cells/mL were transferred to a fresh tube containing differentiation medium (ESC medium without LIF and 2 inhibitors). Drops of the cell suspension were spotted on the lid of a dish containing 1x PBS (phosphate-buffered saline: 2.7 mM KCl, 137 mM NaCl, 1.8 mM KH2PO4, 10 mM Na2HPO4, in ddH2O, pH ~ 6.8) on the bottom. After 3 days, EBs were transferred to a new dish containing a differentiation medium and incubated for 5 days. 24 hours before staining, EBs were trypsinized and seeded on gelatinized coverslips.

HEK-293 human cell line (CVCL_6974) was obtained from Invitrogen (Paisley, UK). HeLa cells were obtained from Stuart Orkin´s Laboratory (HMS, Boston, USA). Bj-hTERT cell line was obtained from Dr. Mathias Rosenfeldt, University of Wuerzburg, Germany. HEK-293 and HeLa cells were transfected using polyethyleneimine (PEI) (pH 7.0, Cat. No.: 40827–7, Sigma-Aldrich Chemie GmbH, Steinheim, Germany) as previously described [48]. For 10 cm diameter dishes, 90 µL PEI and 30 µg plasmid DNA were added to 900 µL DMEM without supplements and mixed by vortexing. 12 µL PEI and 4 µg plasmid DNA were added to 200 µL DMEM for transfection of 6-well plates (amount per well). The mixtures of PEI and DNA were combined, vortexed for 80 seconds, and incubated at room temperature for 30 minutes to 1 hour. Then, the mixture was added dropwise to the cells growing to a confluency of 80% at the moment of transfection. For transfection by electroporation, MTF wild type cells (obtained from Bird Laboratory, Edinburgh, UK) and BHK clone 2 cells (obtained from Spector's Laboratory, CSHL, USA) were transfected using an AMAXA Nucleofector® system II (Lonza, S/N: 10700731) using a self-made buffer (5 mM KCl (Sigma-Aldrich Cat.No.: 7447–40–7), 15

mM MgCl2 (Sigma-Aldrich Cat.No.: 7786–30–3), 120 mM Na2HPO4/NaH2PO4 (Sigma-Aldrich Cat.No.: 7558–79–4) pH 7.2, 50 mM Mannitol (Caesar & Loretz, Cat.No.: 69–65–8)) [49] with default programs for each cell line. Cells growing at 60–80% confluency were trypsinized and centrifuged (6.5 min; 1,400 rpm) the day before the experiment. Afterward, cells were resuspended in 100 µL of nucleofection buffer containing 5 µg of the plasmid DNA. Subsequently, cells were seeded on gelatinized coverslips or glass-bottom p35 plates for live-cell experiments (rF3H). For transfection of mouse embryonic stem cells (mESC), the Neon electroporation system was used (ThermoFisher Scientific) following the manufacturer's instructions.

All cell lines' characteristics, sources, and references are listed in Supplementary Table S2.

Mass spectrometry sample preparation, LC-MS/MS, and data analysis

GFP-tagged TET proteins were expressed in HEK293T cells. Cell lysis and immunoprecipitation with the GFP-Trap (ChromoTek GmbH, Martinsried, Germany) were performed as described previously [50]. After co-immunoprecipitation, protein samples were digested on beads with trypsin according to standard protocols. Wash buffer (20 mM Tris-HCl (pH 7.5), 300 mM NaCl, and 0.5 mM EDTA) was used to rinse the samples on beads two times, followed by two washes with immunoprecipitation buffer (20 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 0.5 mM EDTA). Sample preparation and mass spectrometry were performed as described in [51]. For data analysis, all raw files were analyzed using the MaxQuant computational proteomics platform (version 1.4.1.6) [52] as described in [51]. Peak lists were searched against the UniProt database (https://www.uniprot.org/) with an initial mass deviation of 7 ppm and fragment ion deviation of 0.5 Thomson using carbamidomethylation as a fixed modification. In addition, oxidation of methionine and acetylation of the N terminus were used as variable modifications. For protein quantification, all unmodified and oxidized methionine- and N-acetylation-containing peptides were used with the label-free quantitation algorithm. Perseus software (version 1.5.0.15) [52] was used to analyze MaxQuant output data. Statistical significance was tested using a Student's two-tailed paired t-test. Experiments were performed in biological triplicates.

Structural modeling with AlphaFold 3 and AlphaFold 2-Multimer

The AlphaFold3 web interface was used to screen protein-protein-RNA interactions (https://alphafoldserver.com/) [53]. To rank the predictions, both pTM and ipTM scores were used. The predicted template modeling score (pTM) and the interface-predicted template modeling (ipTM) are measures of the accuracy of the entire structure [54,55]. As a guide, pTM scores above 0.5 indicate that the overall predicted fold for the complex might be similar to the true structure. Complementary to pTMs, the ipTM score measures the interface, and the accuracy of the predicted relative positions of the subunits within the complex. For ipTM scores, values higher than 0.8 represent high-quality predictions with high confidence, while values below 0.6 suggest a likely failed prediction. PAE and pLDDT were also taken into consideration after a close inspection of the models. PAE plots for full-length protein predictions and preexisting models in the 3D structures databases (AlphaFoldDB) were used to perform sequence fragmentation for further predictions. Well-predicted domains (pLDDT > 90) were not fragmented and were taken as a unit core. As an RNA sequence in protein-protein-RNA interactions, we selected a short fragment located in the intron-exon limit of the spliceable firefly luciferase (5'-CCCACAGCCGGCG-3') (splicing reporter system, pc3336). Protein fragments were designed using available monomeric structural models from AlphaFold as provided by the AlphaFold database [56]. To model TET1-DNA interaction, we selected a short DNA sequence in the LINE 1 5' UTR (5'-GCGCA CCTTC CCTGT AAGAG AGCTT GCCAG CAGAG AGTGC TCTGA-3') that is hydroxymethylated by TET1s [57]. To

model TET1-RNA interaction, we used the following fragment of the splicing reporter pc3336: 5'-AAAACAUAAAGAAAGGCGUGAGUCUAUGGGA-3.' A local installation of AlphaFold-Multimer 2.3.2 was used to perform selected structural modeling with AMBER relaxation [58]. Five predictions were generated per model, and the prediction with the highest model confidence was used for verification of AlphaFold 3 predictions. All protein sequences were extracted from UniProt [59]. Data analysis and plotting were done with Python (pandas and numpy packages). Protein structure images were generated with ChimeraX [60,61]. All the structural models, generated '.pbd' files, and associated data, including PAE values, '.json,' and scores, are deposited and available at the repository TUdatalib (https://doi.org/10.48328/tudatalib-1657.3).

Immunostaining of mammalian cells

Immunostainings of cells were performed as described in [62] and [57]. First, cells were seeded on gelatine-coated glass coverslips at least 24 hours before fixation. Then, cells were washed with 1x PBS once and fixed with 3.7% formaldehyde (Sigma-Aldrich Chemie GmbH, Steinheim, Germany, Cat. No.: F8775) in 1x PBS for 10–15 minutes. After fixation and one washing step with PBS-T (1x PBS, 0.01% Tween-20), cells were permeabilized with 0.5% Triton X-100 in 1x PBS for 20 minutes. Next, cells were washed again with 1x PBS and blocked in 0.02% fish skin gelatine or 1% bovine serum albumin (BSA) in 1x PBS for 30 minutes to 1 hour. Following blocking, the primary antibodies were diluted in blocking solution and incubated for 70 minutes at 37°C in a humid chamber, after which three washing steps with 0.01% Tween 20 in PBS (10 minutes each) were performed. Incubation with the secondary antibodies diluted in blocking solution (1% BSA in PBS) was performed for 1 hour in the dark. After secondary antibody incubation, three washing steps with 0.01% Tween in PBS for 10 minutes each were performed. Finally, DNA was counterstained with DAPI (4,6-diamidino-2-phenylindole, 10 g/ml, Cat. No.: D27802, Sigma-Aldrich Chemie GmbH, Steinheim, Germany) for 10 minutes, washed with 1xPBS, and samples were mounted in Mowiol 4–88 (Cat. No.: 81381, Sigma-Aldrich Chemie GmbH, Steinheim, Germany) containing 2.5% DABCO (1,4-diazabicyclo[2.2.2]octane, Cat. No.: D27802, Sigma-Aldrich Chemie GmbH, Steinheim, Germany) after a previous short immersion in ddH2O.

For immunostaining of mouse tissues, CrK $(SO_4)_2$ coated slides with paraffin-embedded mouse tissue sections were incubated for 3 hours at 65°C. To remove paraffin, 100% xylol (3x 5 minutes) and a series of decreasing ethanol concentrations (100%, 96%, 90%, 80%, 70%; each for 5 minutes) were used. For antigen retrieval, slides were autoclaved (100°C for 30 minutes, end temperature 80°C) in 1 mM Tris-EDTA (pH 8.5) with subsequent incubation in 0.05% Tween in PBS two times for 15 minutes and briefly in PBS. Paraffin-embedded tissues were blocked with 1% BSA in PBS for 30 minutes. After blocking, incubation with primary antibodies was performed overnight at 4°C, followed by washing as described above. Secondary antibodies were incubated in 1% BSA for 1 hour. After antibody incubation, samples were washed with 0.01% Tween in PBS (0, 5, 15 minutes). DNA was counterstained using DAPI, with a subsequent washing step using PBS and ddH_2O. Slides were mounted in Mowiol.

Cryosections on microscopy slides (SuperFrost Ultra Plus, Roth, Germany) were dried up for 30 minutes and transferred to sodium citrate buffer 10 mM (pH 6.5) for 3 minutes. For antigen retrieval, slides were heated up to 80°C (water bath) in the same solution, followed by incubation in 0.5% Triton X-100 in PBS for 1 hour. Both primary and secondary antibody incubations were performed in a blocking solution containing 20 µg/mL DAPI. Incubations were performed in humid dark chambers overnight [63]. Wash steps were performed using 0.01% Triton X-100 in PBS three times for 30 minutes at 37°C. Slides were mounted in Vectashield.

For RNase treatments followed by extraction and immunofluorescence, MTF cells were seeded on glass coverslips. 24 hours later, cells were washed once with Triton 0.01% in PBS and incubated for 7 minutes with RNaseA 10 mg/ml in 1x PBS. After RNaseA treatment, cells were fixed with

3.7% paraformaldehyde in 1x PBS and immunostained with antibodies against TET1 and SC35. Subsequent steps were performed as described before. As treatment control, cells subjected to the same treatments (extraction and RNaseA incubation) were stained with propidium iodide (Cat. No.: P4170, Sigma-Aldrich) 1:200 in 1x PBS, and the respective signal was quantified by high-content microscopy ($n > 1000$).

All primary and secondary antibody details, as well as dilutions used, are available in Supplementary Table S3.

Immunostaining of Drosophila salivary gland cells

For the analysis of colocalization of SC-35 and RNA Pol II in the salivary gland of the larvae of Drosophila, tissues were dissected in PBS, fixed in 4% paraformaldehyde in PBS for 20 minutes at RT, and washed in PBST (PBS with BSA and 0.3% Triton X-100). Antibody in-situ stainings were done as described previously [64] using rabbit anti-HA (1:1000; Sigma), mouse anti-SC35 (Sigma-Aldrich S4045 ascites, 1:1000), rat anti-RNA Pol II (anti-Ser 2, clone 3E10, 04–1571 Merck-Millipore, 1:1000) and visualized with Alexa Fluor 488 (1:250; Molecular Probes), Alexa Fluor 546 (1:250; Molecular Probes) or Alexa Fluor 647 (1:250; Molecular Probes). To perform microscopy imaging, the tissues were mounted using Vectashield (Vector Labs). Slices were visualized with a confocal microscope Leica TCS SP8. Microscopy images were processed using Fiji (https://Fiji.nih.gov/ij/).

Recombineering was used to clone the genomic part of the core TET gene into pUC 3GLA UAS from a genomic BAC clone as described [65,66] by adding a HA Tev FLAG C-terminal tag. The N-terminal part for each isoform was then added from PCR-amplified cDNA. For the catalytically dead mutant short isoform, the conserved HxD iron binding motif required for the catalytic activity of TET/AlkB dioxygenase family enzymes was mutated to AxA [33]. Transgenic lines were generated by phiC31-mediated insertion into the attP2 landing site. For immunohistochemistry, the GFP marker was removed by Cre recombinase.

RNA-seq analysis

To analyze alternative splicing events and differential gene expression in Drosophila TETnull mutants in comparison with wild-type, the approach used in [67,68] was performed, using published RNAseq data [67,68]. RNA-seq data sets can be found in GEO under the accession number GSE116212. GO analysis was done with STRING using the rank list enrichment analysis.

Microscopy and image analysis

A Nikon Eclipse Ti microscope controlled by the software Volocity 6.3 and equipped with an UltraView VoX spinning disk system (Perkin Elmer; Waltham, MA) and a Hamamatsu C9100–50 cooled EMCCD14 camera was used to acquire confocal images. The objective used was a Nikon CFI Plan Apo VC 60x oil immersion (60/1.44 NA), and lasers with excitation lines 405, 488, 561, and 640 nm were used. Z-stack images were acquired with a voxel size of XY = 120 nm and a Z-step length of 0.3 µm. In addition, the confocal microscope Leica TCS SPE-II and SP5 systems were used for image acquisition, mounted on a DMi8 stand and controlled by the Leica LAS X software. In these microscopes, a Leica ACS APO 20x/0.60 NA corrected for oil immersion and Leica ACS APO 63x/1.30 NA Oil CS 0.17/E,0.16 were used as objectives, in combination with the laser with 405, 488, 561, and 635 nm. The images of mouse tissues were acquired using a Leica TCS SP5 confocal microscope (Milton Keynes, UK) equipped with Plan Apo 63x/1.40 NA oil immersion objectives and lasers with excitation lines 405, 488, and 633 nm. Computational image analysis was performed with Fiji software (http://fiji.sc/). For imaging processing, the 'Gaussian Blur' filter with a sigma radius of 1.0 was applied to all images. Detailed colocalization analysis was performed by implementing the Python-based image analysis platform Priithon (http://code.google.com/p/priithon/), which was used to calculate the H-coefficient for the different conditions as in [69]. Accumulation analysis and rF3H (RNA-fluorescence 3-hybrid) assay image analysis were performed as described in [62]. To image and quantify splicing reporter assays, the

Operetta high-content screening system (Perkin Elmer, UK) was used in wide-field mode, equipped with a Xenon fiber optic light source and a 20×/0.45 NA long working distance or a 40×/0.95 NA objective. For excitation and emission, the following filter combinations were used: 360–400 nm and 410–480 nm for DAPI, 460–490 nm and 500–550 nm for Alexa-488, as well as 560–580 nm and 590–640 nm for Alexa-594. Fluorescence intensity levels were quantified with the Harmony software (Version 3.5.1, PerkinElmer, UK). All imaging systems used are summarized in Supplementary Table S4.

RNA fluorescent three-hybrid assay (rF3H)

We performed the RNA fluorescent three-hybrid assay to address the interaction between TET1 proteins and RNA, specifically an mRNA mimic [70]. To this end, BHK cells were transfected with plasmids encoding an RNA trap fused to EGFP and the LacI protein for targeting the LacO present in this cell line (pc4555), an mRNA mimic (pms2-PABCP1-mCherry, pc4573), and pmCherry-TET1-CD (pc2547). As a positive control of the assay, cells were transfected with the RNA trap, the mRNA mimic, and PABPC1-mCherry, which was shown to interact with the mRNA mimic [70]. As a negative control, only the mRNA trap plus PABPC1-mCherry were used for transfection. Cells were imaged live 8–12 h post-transfection and confocal Z-stacks (voxel size, 0.12 × 0.12 × 0.5 µm) were acquired using the confocal microscope Leica TCS SP5 II. Z-stacks were analyzed and mounted using Fiji (https://Fiji.nih.gov/ij/). Cells showing GFP and mCherry signals were selected. The LacO was segmented in each nucleus for image analysis using the GFP signal and generating an ROI (region of interest). Then, sum intensities for mCherry proteins were measured in this ROI. A second ROI of the same shape and size was used to measure the sum intensity of mCherry proteins in a region outside the LacO. The relative accumulation at the LacO was calculated as the ratio between mCherry sum intensity in the LacO and mCherry sum intensity outside the LacO.

Luciferase and NMD splicing reporter assay

For luciferase splicing reporter assay, HEK cells were transfected with PEI using 2 µg of mCherry-tagged TET1 constructs (TET1-CD/TET1-CDmut) or mCherry alone, in combination with 2 µg of plasmid DNA encoding firefly luciferase constructs with either one (pc3336) or two introns (pc3337). 24 hours post-transfection, the cell culture medium was removed, and the cells were washed with 1x PBS. 200 µL of Cell Lysis Reagent (Luciferase Assay System, Promega) was added and incubated for 10 minutes at room temperature. Cells were scraped and transferred to a new 1.5 mL tube. 50 µL of the cell extract was mixed with 100 µL of Luciferase Assay Reagent (Luciferase Assay System, Promega) in a 96-well plate, and luminescence was measured immediately for 20 seconds per well using a microplate reader (Infinite® 200 PRO series, TECAN). Afterward, the fluorescence signal intensity of mCherry was measured for the whole plate and used for normalization of the luminescence signal. For NMD reporter assay with EGFP, J1 mESC wild type and TET or Dnmt triple knockout cell lines were transfected with N-terminal mCherry-tagged catalytic domains of TET1, TET2, and TET3. mCherry alone was used as a control. In this assay, the mRNA is translated when the exon 8 at RNP-DL is spliced out. If no splicing occurs, the mRNA is degraded, and no fluorescence signal is detected. The intensity of the EGFP signal was quantified by microscopy and image analysis using the Operetta high-content microscopy system (Perkin Elmer, UK) in wide-field mode, equipped with a Xenon fiber optic light source and a 20x/0.45 NA long working distance objective. For excitation and emission, the following filter combinations were used: 360–400 nm and 410–480 nm for DAPI, 460–490 nm for GFP, and 560–580 nm for TexasRed (mCherry). Fluorescence intensity levels were quantified with the Harmony software (Version 3.5.1, PerkinElmer, UK)

In vitro splicing assay

In vitro splicing experiments were performed with MINX-M3 pre-mRNA [71–73]. DNA templates

for run-off IVT (in vitro transcription) reactions were generated via digesting the plasmid with XbaI restriction enzyme. MINX-M3 pre-mRNAs with different methylation status of cytosines were in vitro transcribed in 50 µl reactions containing T7 transcription buffer (5X stock: 600 mM HEPES-KOH pH 7.5, 160 mM MgCl2, 10 mM spermidine, 200 mM DTT (Dithiothreitol)); 7.5 mM ATP; 7.5 mM CTP, mCTP, or hmCTP; 1.5 mM GTP; 1.5 mM UTP; 5 mM 3mGpppG cap analog; 5 µl 32P-UTP (3000 Ci/nmol, 10 µCi/µL); 50 ng/µL template DNA; 6 µL T7 polymerase, 2 µL RNasin and 0.5 µL YIPP IVTs were performed for 2 h and 37°C, and RNA was subsequently purified via 4% denaturing PAGE and extracted with phenol-chloroform. In vitro splicing reactions with IVT pre-mRNAs were performed for 0–90 minutes at 30°C in the presence of 40% HeLa nuclear extract, 3 mM $MgCl_2$, 65 mM KCL, 20 mM HEPES-KOH pH 7.9, 2 mM ATP, and 20 mM creatine phosphate. RNA was extracted from splicing reactions with protein K digestion followed by phenol-chloroform extraction and ethanol precipitation. Isolated RNA was then analyzed on 14% denaturing PAGE and visualized with autoradiography.

Co-immunoprecipitation experiments and Western blotting

To analyze the protein–protein interaction between TET and splicing factors, we performed co-immunoprecipitation experiments as described in [62,74]. First, HEK cells were co-transfected with TETs and U2AF1/U2AF2/SC35 expression vectors using PEI and harvested by trypsinization 48 hours after transfection. Cells were trypsinized and centrifuged for 10 minutes at 2,000 rpm and 4°C. The pellets were resuspended in 200 µL ice-cold lysis buffer containing 20 mM Tris HCl (pH 8), 150 mM NaCl, 1.5 mM MgCl2, 0.4% NP-40, 0.2 mM EDTA, and protease inhibitors 1 mM AEBSF (4-(2-Aminoethyl) benzyl sulfonyl fluoride hydrochloride, Cat. No.:A1421.0100, VWR, Radnor, PA, USA), 1 mM E64 (Cat. No.: E3132, Sigma-Aldrich, St Louis, MO, USA), 1 nM Pepstatin A (Cat. No.: 77170, Sigma-Aldrich, St Louis, MO, USA), PMSF (10 µM, Sigma-Aldrich, St. Louis, MO, USA/Solarbio; catalog #P8340) and AEBSF (1 mM, AppliChem, Darmstadt, Germany). For lysis, cells were homogenized with a syringe using 21 G needles, and 20 strokes per sample. Cell pellets were maintained on ice during homogenization and incubated on ice afterward. After 30 minutes in ice (with repeated vortexing every 10 minutes), the cell lysates were cleared by centrifugation (15 minutes at 13,000 ×g and 4 °C). For the input fractions, 15% of the lysate (approximately 30 microliters) was used. For the binding fractions, the rest of the lysate was incubated with GFP-binder beads [75]. Incubation with the beads was performed on rotation at 4°C for at least 90 minutes in buffer with 3% PMSF and 1 mM Ribonucleoside Vanadyl Complex RNase Inhibitor (NEB). After incubation of the protein lysate with the beads, they were washed three times to remove non-bound proteins with 500 µL washing buffer containing 20 mM Tris HCl, 150 mM NaCl (pH 8), 1.5 mM MgCl2, and 0.2 mM EDTA. Washing was performed by centrifugation of the samples at 2000 rpm, followed by supernatant removal. After washing and resuspension in a small volume of loading buffer 4x SDS (400 mM DTT, 200 mM Tris/HCl pH 6.8, 8% SDS, 0.4% bromophenol blue and 40% glycerol), input and bound fractions were boiled at 95°C and separated on 8% SDS-PA (sodium dodecyl sulfate – polyacrylamide) gels.

The SDS-PAGE and Western blotting experiments were performed as in [76]. 4 microliters of protein ladder maker were loaded into the polyacrylamide gel (Color Prestained Protein Standard, Broad Range 10–250 kDa, NEB, catalog number P7719L; MWP06 BlueEasy Prestained Protein Marker, Nippongenetics). Briefly, transference of co-immunoprecipitated samples from the gel to the nitrocellulose membrane (GE Healthcare, München, Germany) was performed in a semi-dry blotting chamber (Bio-Rad Laboratories) for 50 minutes at 25 V. Ponceau S staining was used after this time to verify the successful transfer onto the membranes. After colorimetric imaging of the membranes stained with Ponceau S, these were briefly washed with ddH2O and subjected to blocking for 1 hour in low-fat milk 3% dissolved in 1x PBS at room temperature on agitation. After blocking, membranes were incubated with primary antibodies

diluted in a blocking buffer. This incubation was performed overnight at 4°C in rotation (with membranes placed inside 50 mL Falcon tubes). Following primary antibody incubation, the membranes were washed 3 times for 10 minutes each with 1x PBS + 0.02% Tween-20. After this, incubation with the secondary antibodies was performed for 1 hour, followed by washing as described before. For visualization of the bands, horseradish peroxidase (HRP) conjugated secondary antibodies were used. All the characteristics and dilutions of primary and secondary antibodies and dilutions used are described in Supplementary Table S3. To develop the membranes, RNeasy® ECL Western Blotting Substrate was used (Cat. No.: 32209, ThermoFisher Scientific, Waltham, MA, USA). The Amersham AI600 Imager with a CCD camera (GE Healthcare, Chicago, II, USA) was used to image immunoreactive bands and Ponceau S staining. Incremental exposition times were used in all cases. For a better composition of the figures and due to space restrictions, cutouts of the membranes were made. Uncropped blots showing the full range of the protein ladder are available in Supplementary Figure S2, together with replicates. Unprocessed scans for all the blots are provided with the data sets uploaded to TUDatalib (https://doi.org/10.48328/tudatalib-1657.3)

RNA isolation and slot blotting

To investigate the hm5C amount in J1 ESCs, total RNA was isolated using a Qiagen RNeasy® Kit. Isolation was mainly performed following the manufacturer's instructions with the following changes. Cell pellets were vortexed with 350 µL RLT buffer and passed 10 times through a blunt 20 G needle fitted on an RNase-free syringe. 350 µL of 70% ethanol (in DEPC water) was added and mixed by pipetting. 700 µL were transferred to an RNeasy® spin column and centrifuged for 15 seconds at 10,000 rpm. At the end of the protocol, RNA was eluted twice with 30 µL RNase-free water. After RNA isolation, DNA was digested using rDNase from Nucleospin® Triprep Kit (Macherey Nagel). For RNA clean-up, samples were filled up to 100 µL and mixed with 350 µL RLT buffer containing 2-mercaptoethanol (1:100). 250 µL of 100% ethanol was added and mixed by pipetting. 700 µL were transferred to an RNeasy® column (Qiagen) and centrifuged for 15 seconds at 10,000 rpm. Afterward, 500 µL RPE buffer was added and samples were centrifuged again. After discarding the flowthrough, 500 µL 80% ethanol (in DEPC water) was added, and samples were centrifuged for 2 minutes at 10,000 rpm. Columns were placed into fresh tubes and RNA was eluted with 20 µL RNase-free water by centrifugation for 1 minute at 13,200 rpm. To assess the concentration and purity of RNA, the ratio of absorbance at 260 and 280 nm was measured on a TECAN infinite M200 plate reader (Tecan Group Ltd., Maennedorf, Switzerland).

For slot blot analysis of hm5C in total RNA, isolated samples were blotted on nylon membranes pre-equilibrated with 20x SSC. For controls, plasmid DNA was digested with HindIII and EcoRI (20 µL total volume, NEB buffer 2.1) for 1 hour at 37°C, and an RNA sample was digested with RNase A (20 µL/mL) for 30 minutes at 37°C. Respective RNA amounts were applied in duplicates to the slots of a blotter (Schleicher & Schuell) and pulled through by vacuum. After air drying, the membrane was UV-crosslinked with 120,000 µJ/cm2 two times for 1 minute. To calculate the amount of membrane-bound RNA, half of the samples were stained with 0.02% methylene blue in 0.3 M sodium acetate. The other half was blocked with 3% milk in PBS for 30 minutes. The 5hmC antibody was incubated overnight at 4°C, and the secondary antibody for 1 hour at room temperature. Following antibody incubations, the membrane was washed with 0.1% Tween in PBS (3x 10 minutes). To detect the chemiluminescent signal, a Luminol Solution (Thermo Scientific, Pierce ECL Plus Western Blotting Substrate) was applied to the membrane, and images were acquired with an Amersham Imager 600 (Amersham) (Supplementary Table S4: Imaging systems).

Data visualization and statistics

For this study, RStudio (versions V1.2.5033 and V2023.03.1–446, https://rstudio.com/) and Microsoft® Excel® for Mac 2011 (Version 14.7.7) were used for data visualization, analysis, statistics,

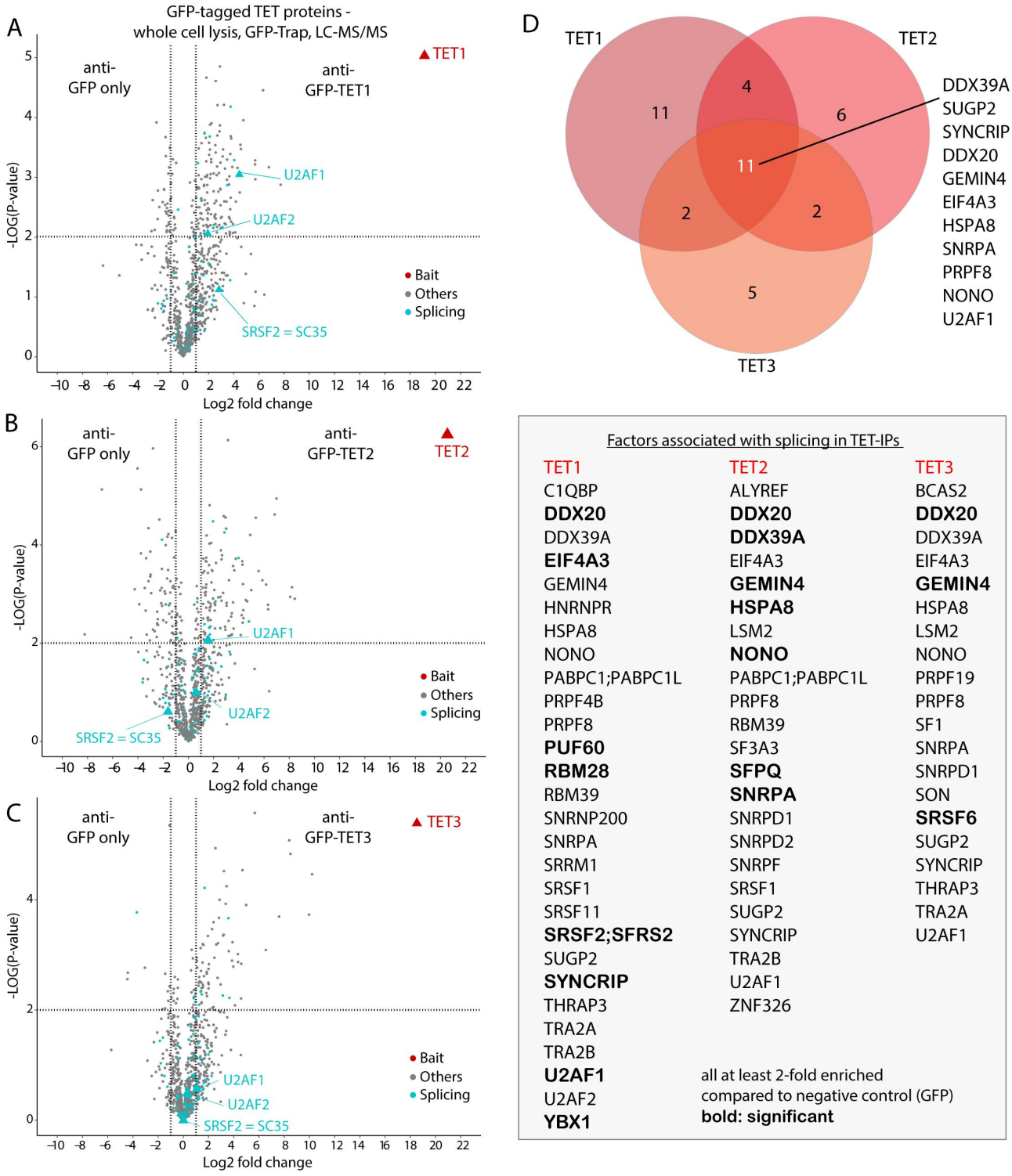
and plotting (unless stated otherwise). In all barplots, the average value of the distribution is shown, with the whiskers representing the standard deviation (95% confidence interval). For graphs showing boxplots, in all cases, the box shows the distribution of 50% of the data. This range starts in the first quartile (25%) and ends in the third (75%), with the line inside the box representing the median value and with whiskers representing the upper and lower quartiles not included in the box. Outliers are defined as 1.5 times the interquartile range and are excluded from most of the plots. Statistical comparisons were applied when considered necessary using an independent two-group comparison test: Wilcoxon-Mann-Whitney or One-Way ANOVA. Related to statistical significance, not significant (n.s.), is assigned to p-values > or equal to 0.05; one star (*) is assigned to p-values < 0.05 and > or equal to 0.005; two stars (**) is assigned to values < 0.005 and > or equal to 0.0005; three stars (***) is given for values < 0.0005; This information is located in the plots between the top of two samples subjected to comparison. Statistical values (number (#) of cells (N), mean, median, standard deviation (SD), and p-values are summarized in Supplementary Table S5.

Results and discussion

TET proteins interact with the splicing factors U2AF1 and U2AF2

In the course of mass spectrometry experiments investigating TET proteins interactome and post-translational modifications [51], we found a significant enrichment of splicing factors in the proteins detected by LC-MS/MS. In this analysis, GFP-tagged murine TET proteins or GFP as a control (Supplementary Table S1) were expressed in human HEK293T cells. GFP-tagged proteins were pulled down from whole cell lysates with GFP-binder beads and subjected to LC-MS/MS (Figure 1). An overall sequence coverage of ~50% was achieved for TET1, ~60% for TET2, and ~65% for TET3 (Supplementary Data S1). Volcano plots show the analysis of these experiments for TET1 (Figure 1A), TET2 (Figure 1B), and TET3 (Figure 1C), where each circle/triangle represents a protein detected by LC-MS/MS and the X-axis depicts differences in protein abundance in the respective pulldowns as Log2 fold change (GFP-TET versus GFP control). Factors more enriched in TET-IP (immunoprecipitation) have positive Log2 fold change values (label-free quantification by MaxQuant). Mass spectrometry analysis of TET-associated factors showed stronger enrichment for the splicing factors U2AF1, U2AF2, and SRSF2 (also known as SC35) (blue triangles) for TET1 pull-downs, while this enrichment was less significant for TET2 and TET3. The Venn diagram in Figure 1D shows splicing-associated factors that were found to interact with all three TET proteins and illustrates the overlap of splicing factors identified in all TET pull-downs. Again, TET3 and TET2 showed less enrichment of splicing-associated factors (6 and 5, respectively) compared to TET1 (11 splicing factors). Nonetheless, there was a shared set of splicing factors interacting with all three proteins. Specifically, the proteins U2AF1, U2AF2, and SC35 were significantly enriched in the TET1 and TET2 pull-down. These proteins are critical components of the spliceosome and are involved in pre-mRNA processing, which suggests a functional interaction between TET dioxygenase enzymes and splicing machinery. Given the role of TET proteins in DNA demethylation and potential involvement in RNA modifications (e.g., hm5C in RNA), these interactions could reflect a regulatory role in gene expression via splicing.

Mass spectrometry studies provided insights into both shared and specific protein interactions for TET proteins, suggesting that a core set of splicing factors may be associated with TETs, while unique interactions may point to specific functions for each TET protein in different cellular contexts. Spliceosome assembly occurs by ordered interaction of the spliceosomal snRNPs (small nuclear ribonucleoproteins) and numerous other splicing factors. During exon definition, the U1 snRNP binds to the 5'ss downstream of an exon and promotes the association of U2AF (U2AF1 and U2AF2) with the polypyrimidine tract/3'ss upstream of it. Splicing enhancer sequences within the exon (ESEs) recruit proteins of the SR protein family (among them SC35), which establish a network of protein–protein interactions between the different components of the spliceosome



Factors associated with splicing in TET-IPs

TET1	TET2	TET3
C1QBP	ALYREF	BCAS2
DDX20	**DDX20**	**DDX20**
DDX39A	**DDX39A**	DDX39A
EIF4A3	EIF4A3	EIF4A3
GEMIN4	**GEMIN4**	**GEMIN4**
HNRNPR	**HSPA8**	HSPA8
HSPA8	LSM2	LSM2
NONO	**NONO**	NONO
PABPC1;PABPC1L	PABPC1;PABPC1L	PRPF19
PRPF4B	PRPF8	PRPF8
PRPF8	RBM39	SF1
PUF60	SF3A3	SNRPA
RBM28	**SFPQ**	SNRPD1
RBM39	**SNRPA**	SON
SNRNP200	SNRPD1	**SRSF6**
SNRPA	SNRPD2	SUGP2
SRRM1	SNRPF	SYNCRIP
SRSF1	SRSF1	THRAP3
SRSF11	SUGP2	TRA2A
SRSF2;SFRS2	SYNCRIP	U2AF1
SUGP2	TRA2B	
SYNCRIP	U2AF1	
THRAP3	ZNF326	
TRA2A		
TRA2B		
U2AF1		
U2AF2		
YBX1		

all at least 2-fold enriched
compared to negative control (GFP)
bold: significant

Figure 1. Mass spectrometry analysis reveals novel interactions between TET proteins and splicing factors. (A) TET1, (B) TET2, and (C) TET3 mass spectrometry analysis of TET-associated factors. GFP-tagged murine TET proteins or GFP as control were expressed in human HEK293T cells. GFP-tagged proteins were pulled down from whole cell lysates with the GFP-binder beads and subjected to LC-MS/MS. Volcano plots: each circle/triangle represents a protein detected by LC-MS/MS; X-axis depicts differences in protein abundance in the respective pulldowns as log2 Fold change (GFP-TET versus GFP control); factors more enriched in TET-IP have positive Log2 Fold change values (label-free quantification by MaxQuant) Y-axis depicts the negative log10 of the p-value of a student's t-test (triplicate samples) dashed black lines: significance border (FDR = 0.05, S0 = 2), blue circles/triangles denote splicing factors. A higher value corresponds to a lower p-value, indicating stronger statistical significance. Proteins outside the dashed lines (upper right or left corners) are significantly enriched or depleted, respectively. (D) Venn diagram shows splicing-associated factors that were found to interact with different TET proteins. TET3 and TET2 showed slightly less enrichment of splicing-associated factors compared to TET1, as shown in the list below.

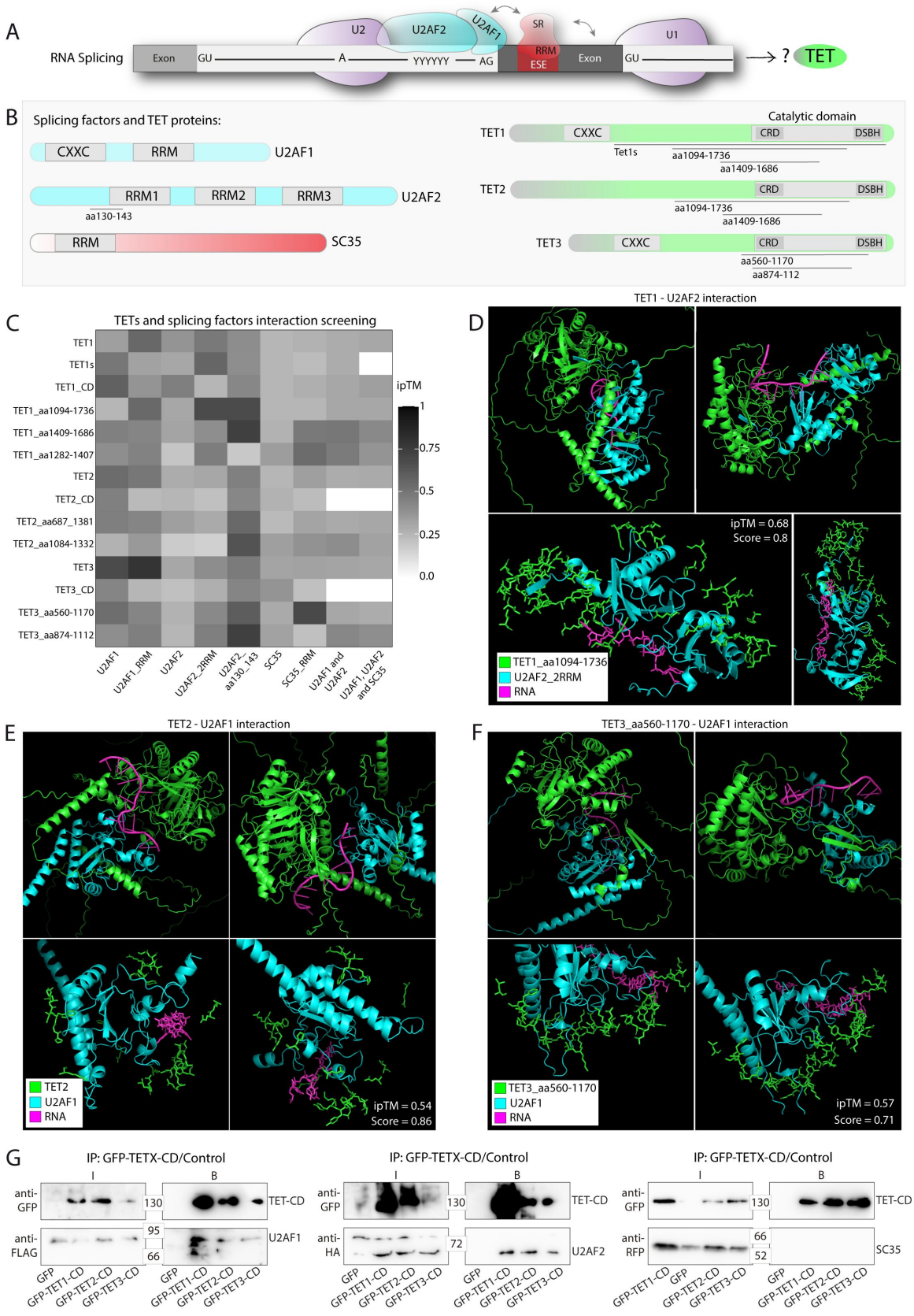


Figure 2. Screening and validation of TET protein-protein-RNA interactions by structural modeling with AlphaFold 3. (A) the diagram shows the process of RNA splicing and the proteins involved, with a pre – mRNA harboring exons and introns (light gray) which are removed during splicing. The splice sites at the start 'GU' and the end 'AG' of the intron are shown, together with the branch point 'A.' The polypyrimidine tract located upstream of the 3' splice site is represented as 'YYYYY.' U1 and U2 snRnps, components of the

(Figure 2A) [77]. To gain more insights into the role of TET proteins in this process, we performed an *in silico* screening for protein-protein-RNA interactions between TET and the splicing factors U2AF1, U2AF2, and SC35 using AlphaFold3 (https://alphafoldserver.com/) [53]. We further expanded this modeling by incorporating a short RNA sequence within the intron/exon limit. Previous studies reported significant decreases in the sensitivity of AlphaFold for complex structure predictions when using full-length sequences [78]. Therefore, after full-length protein predictions, we also designed and paired protein fragments, either consisting of individual folded or disordered regions in TET proteins and U2AF1, U2AF2, or SC35 subunits. These fragments were chosen based on the PAE (predicted aligned error) values for each pair of residues in the full-length predictions, selecting the protein domains predicted to interact with a low error rate. For most of the structures, these fragments included the cysteine-rich domain in the catalytic domain of TET proteins and the RRM (RNA recognition motif) for the three splicing factors. A diagram of the proteins, including the most relevant domains and the fragments used for AlphaFold predictions, is shown in Figure 2B.

In total, we conducted 121 AlphaFold predictions in our screening for all TET proteins in combination with the different splicing factors and between U2AF and SC35 proteins, all in the presence of the short RNA fragment. The heatmap in Figure 2C shows the ipTMs (interface-predicted template modeling) scores for all these predictions. The ipTM and pTM (predicted template modeling) scores measure the confidence of the overall structure (shown in Fig. S1 and Supplementary Data S2). The pTM is an integrated measure of the prediction, showing the accuracy of the predicted structure for the full complex as a score for the superposition between the predicted structure and the hypothetical true structure. However, the ipTM measures the accuracy of the predicted relative positions of the subunits within the complex, or how they are predicted to interact, being a metric of the predicted interface of the interaction. Therefore, we ranked the resulting models by model confidence using both ipTM and pTM scores. The highest confidence models were selected for each TET protein, with an ipTM of 0.68 for the interaction between TET1_aa1094–1736 and U2AF2_2RRM (Figure 2D), ipTM of 0.54 for TET2 and U2AF1 (Figure 2E), and ipTM of 0.57 for TET3_aa560–1170 and U2AF2 (Figure 2F). The three structural models comprising TET (green), the respective splicing factor (cyan), and RNA (pink) show the layout of the interaction from different angles, with multiple points of predicted interaction between both proteins and the RNA. For better visualization, magnifications of these proximal regions show the U2AF structure and exclusively the residues of TETs located within a range of 5 angstroms (Å). Additional images of the structures and the residue interactions within a range of 4 angstroms (yellow dashed lines) are shown in Fig. S1B-D. Interestingly, a short fragment of U2AF2 (aa130–143) was predicted to interact with TET proteins with the highest scores (0.74 for TET1_aa1409–1686) (Fig. S1B). This fragment, which corresponds to 13 residues, is located

spliceosome, are responsible for recognizing splicing sites. U2AF1 and U2AF2 proteins assist in binding to the 3′ splice sites and branch points. SR proteins (like SC35) bind to exonic splicing enhancers (ESE) to enhance splicing accuracy. (B) schematic representation of U2AF1, U2AF2, and SC35 harboring different RRM domains (RNA recognition motifs), and TET proteins, with their domains, represented: CXXC (zinc finger domain), the conserved catalytic domain in the C-terminus, including the cysteine-rich domain (CRD) and the double-strand beta-helix domain (DSBH), with an insert in between. In addition to the full protein sequences, different fragments of the aminoacid sequences and domains were used to predict protein-protein-RNA interaction, as indicated in the scheme. (C) heatmap showing the ipTM score (interface predicted template modeling score) obtained for AF structure models in the screening for all interactions tested between TET proteins and the splicing factors U2AF1, U2AF2, and SC35. Labels on the ×and y axes indicate the paired protein fragments for structural modeling, including RNA in the prediction. White tiles indicate pairs that were not subjected to structural modeling. (D) the highest-scored structural models were obtained for TET1-U2AF2-RNA, (E) TET2-U2AF1-RNA, and (F) TET3-U2AF predictions. The full structural models are shown from different perspectives (top part), and magnifications are shown at the bottom. In the magnification, only residues within a maximum distance of 5 angstroms to U2AF are shown. (G) Co-immunoprecipitation analysis of TET proteins interaction with splicing factors. The catalytic domain (CD) of TET1, TET2, and TET3, or GFP as a control, were co-immunoprecipitated using GFP-binder beads. Pull-down fractions were analyzed by western blot using anti-FLAG antibody (left, U2AF1), anti-HA antibody (center, U2AF2), or anti-RFP (right, SC35). Inputs 'I' and bound 'B' fractions are shown. Full blot images and replicates are shown in Figure S2. Uncropped and unprocessed images can be found in Supplementary data S5.

in proximity to the second RRM domain in U2AF2 (2RRM) and is not present in U2AF1 or SC35. The results of the interactions screening and structural modeling show that TET1 is predicted to interact mostly with U2AF2, while TET2 and TET3 interactions show higher confidence scores for U2AF1. On the other hand, only TET3 showed an ipTM of 0.7 in the interaction with SC35, but with a lower pTM of 0.59. Higher-ranked predictions were validated using AlphaFold2 (Multimer), obtaining similar score trends and structural models.

Overall, AlphaFold screening supports our initial finding from mass spectrometry, illustrating the structural interplay between TET proteins, splicing factors, and RNA, while also confirming potential differences between the different TET proteins.

Next, we validated AlphaFold3 predictions between TET proteins and splicing factors by co-immunoprecipitation (Figure 2G). To this end, HEK-EBNA cells were transiently co-transfected with plasmids coding for GFP-tagged TET1, TET2, or TET3 catalytic domains (CD) or GFP as control, and U2AF1-FLAG, U2AF2-HA, or mCherry-SC35. GFP-tagged proteins were immunoprecipitated using a GFP-binder [75], and cell lysates were analyzed by western blot. We found that all TET proteins were able to pull down U2AF1 and U2AF2 but not SC35 (Figure 2G and Fig. S2A-B). Additionally, we co-transfected cells with YFP-SC35 and mCherry-tagged TET1, TET2, or TET3 catalytic domains, immunoprecipitated YFP-SC35 as before, and analyzed cell lysates by western blotting. In line with the reverse experiment, immunoprecipitated SC35 was not able to pull down either of the TET proteins (Fig. S2B-right) (Full blot images and replicates are shown in Fig. S2). This again confirmed our AlphaFold protein–protein interaction predictions, showing that U2AF1 and U2AF2 proteins form a multimer complex [10] that interacts directly with the catalytic domain of TET proteins. However, there is no direct interaction between SC35 and TET proteins. These findings support the initial hypothesis based on mass spectrometry data and AlphaFold screening, showing that TET1, TET2, and TET3 interact with the splicing factors U2AF, implicating a potential role in the regulation of RNA splicing.

TET proteins localize at splicing speckles in mouse, human, and Drosophila

Previous studies have shown that RNA serves as a substrate for TET proteins and that hm5C levels increase upon TET overexpression [37]. This evidence supports the role of TET DNA dioxygenases in regulating RNA modifications, which is consistent with our finding that TETs interact with splicing factors. Therefore, we investigated the connection between TET proteins and RNA splicing, investigating whether TET proteins localize to splicing speckles. For these experiments, we focused mostly on TET1 since it showed the highest interaction score with U2AF2, and it is abundantly distributed in a higher number of cell lines and tissues [79]. The splicing factor SC35 is a reliable marker for these subnuclear compartments, and specific antibodies are available. Using immunofluorescence, we addressed a potential colocalization of TET in splicing speckles. We performed immunostainings for both TET1 and SC35 proteins in J1 mouse embryonic stem cells (mESCs) versus differentiated cells (Figure 3A). The localization of TET1 and SC35 in splicing speckles was observed in mESCs and across various differentiation states: embryoid bodies (EB) derived from J1 ESCs, mouse embryonic fibroblasts (MEF), and mouse tail fibroblasts (MTF), using Oct4 as a pluripotency marker.

To validate this observation, colocalization between TET1 and SC35 (gray) and colocalization between TET1 and DAPI (black) were quantified using the H-coefficient (Figure 3A-right). Here, positive colocalization was found for TET1 and SC35 (H-coefficient higher than 1) in all cell lines. However, anti-colocalization (negative values) was found for TET1 and DNA-dense regions stained with DAPI, which are devoid of splicing speckles. As a control for TET1 levels in the different cell lines, we also performed immunofluorescence with TET1-specific antibodies in embryonic stem (wild type and TET TKO), and mouse fibroblast (Fig. S3A). Quantification of nuclear TET1 levels in this immunofluorescence showed similar TET1 levels for wild-type embryonic stem cells and fibroblasts, while TET TKO cells showed TET1 levels resembling the background control (Fig. S3B). From these experiments, we

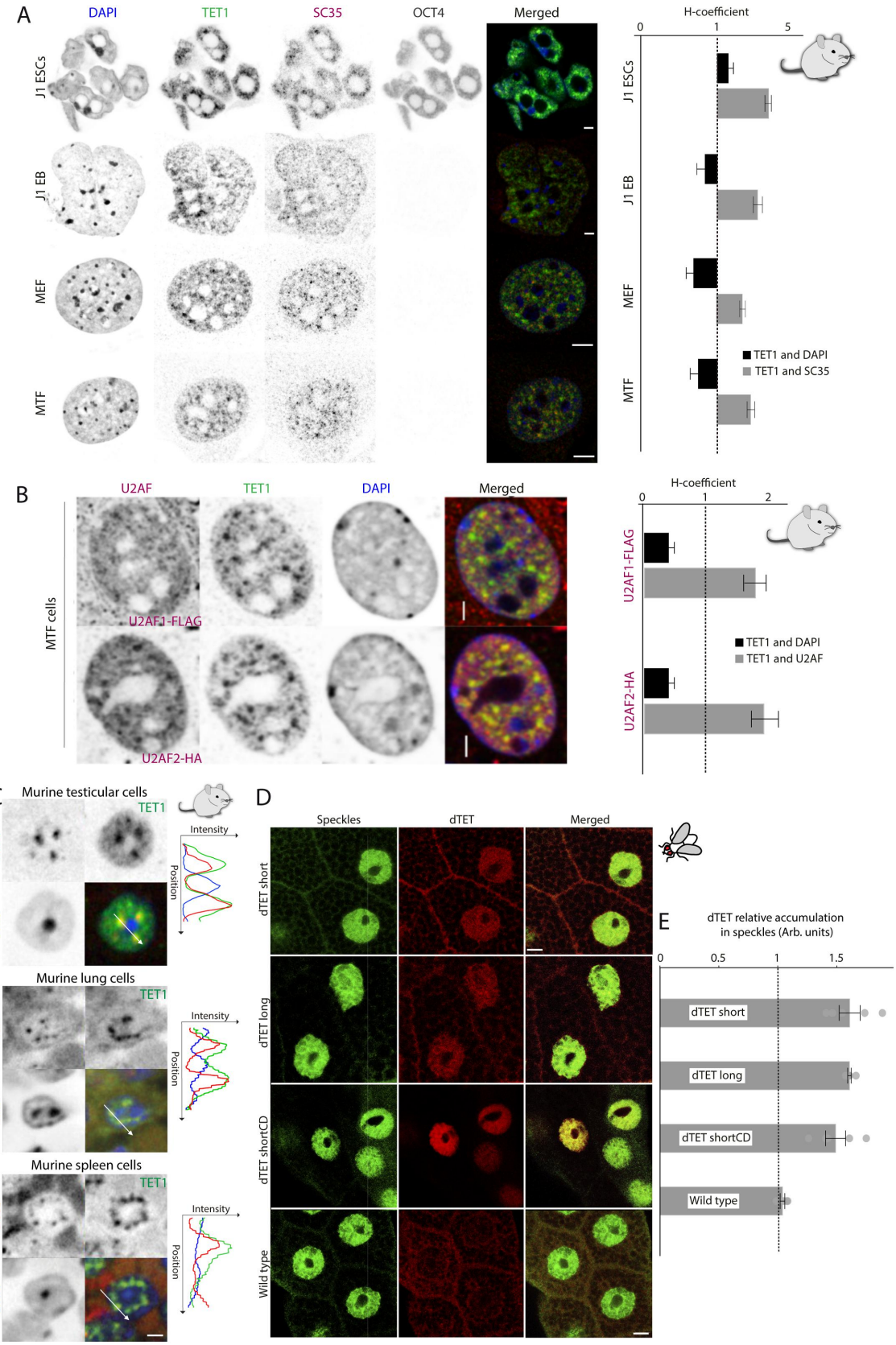


Figure 3. TET1 localizes at splicing speckles, colocalizing with splicing factors in mouse and Drosophila tissues. (A) immunostaining of TET1 and SC35 in various differentiated mouse cell lines with quantified colocalization represented by the H-coefficient. J1 ESCs, embryoid bodies (EB) derived from J1 ESCs, mouse embryonic fibroblasts (MEF), and mouse tail fibroblasts (MTF) were stained for TET1 (green) and SC35 (red). DNA was counterstained with DAPI (blue). To distinguish the undifferentiated state, Oct 4 was used as a

can conclude that TET1 proteins colocalize with splicing speckles in different cell lines and differentiation stages. In addition, similar observations were found performing immunofluorescence for TET2 and TET3 in MTFs (Fig. S4) and human fibroblast Bj-hTERT (Fig. S5), where all three TET proteins were observed to colocalize with SC35 in splicing speckles. Localization of TET1 in the nucleoplasm also agrees with the observation of the Human Protein Atlas (https://www.proteinatlas.org/) in different human cell lines. Next, we performed immunostainings in MTFs transiently transfected with tagged U2AF proteins (U2AF1-Flag and U2AF2-HA) (Figure 3B). MTF cells were chosen for these experiments because they show high endogenous TET1 levels. The quantification of TET1 and U2AF colocalization in speckles also showed H-coefficient values higher than 1 for TET1 and both U2AF proteins and values close to 0 (no colocalization) for dense DNA regions (DAPI-rich) (Figure 3B-right). This confirmed the subnuclear association of TET1 proteins with RNA splicing speckles and splicing factors.

To investigate the nuclear localization of endogenous TET1 in splicing speckles in vivo, we performed TET1 and SC35 immunostainings in mouse tissue sections. The most significant TET1 signals were observed in the testes sections, lung, and spleen (Figure 3C, images of the sections in Figure S6, S7, and S8). TET1 was detected in low-density DNA regions in these immunostainings, clearly colocalizing with SC35-positive splicing speckles. These results show the subnuclear localization of endogenous TET1 proteins at splicing speckles in vivo, pointing to a physiological role of TET1 proteins in regulating gene expression by alternative splicing. On the one hand, TET1 could play a role in chromatin reorganization and decondensation by erasing 5mC, which keeps chromatin in a densely packed state [80]. In line with this hypothesis, TET1s (short isoform) has been shown to cause 5hmC increase and decondensation of heterochromatic regions upon recruitment [62]. On the other hand, TET1 may regulate splicing, leading to the formation of cell-type-specific chromatin reorganization [81]. TET1 localization with SC35 at nuclear speckles may have several physiological consequences. Accordingly, previous studies detected hm5C in RNA of different tissues, like the heart and spleen [37]. These studies, together with our previous results showing TET1 interaction with splicing factors, lead to the prospective role of TET proteins regulating splicing, either via their catalytic activity and m5C oxidation, interaction with splicing factors, and/or non-catalytic functions, some of them discovered recently [62,82,83].

To investigate whether TET's association with splicing speckles was related to DNA or RNA modifications (5mC versus m5C), we selected Drosophila melanogaster as a model. In human and mouse cells, it is not possible to distinguish whether this association is due to TET activity on methylated DNA or RNA. However, previous studies have shown that DNA methylation in Drosophila is nearly undetectable [84–86], while RNA methylation-associated processes are present and functionally important. Drosophila has only one TET gene, which is alternatively spliced to generate two isoforms. While the long isoform with a CXXC zinc-finger DNA binding motif is similar to vertebrate TET1 and TET3, the short

pluripotency marker. The Oct4 channel was omitted for the merge. Colocalization between TET1 and DAPI (black) as well as TET1 and SC35 (dark gray) was quantified using the H coefficient ($n = 10$ cells) and is shown as a barplot on the right-hand side. (B) MTF cells were immunostained for tagged U2AF proteins and endogenous TET1 (green). Ectopically expressed U2AF1-FLAG or U2AF2-HA (red) were visualized using anti-HA or anti-FLAG antibodies. DNA was counterstained with DAPI (blue). The barplot illustrates the H-coefficient ($n = 10$) of TET1 and DNA (black) or U2AF proteins (gray). Scale bar: 5 µm. (C) TET1 localizes to SC35-positive speckles in vivo. Paraffin sections of murine testis, lung, and spleen were immunostained with antibodies against TET1 and SC35. DNA was counterstained with DAPI. Line-profile analysis of selected regions denotes TET1 colocalization with SC35 and anti-colocalization with DNA. (D) colocalization of dTET and nuclear speckles in Drosophila salivary gland cells. Different HA-tagged dTET isoforms (short (row 1), long (row 2), and catalytically dead (cat. dead) short isoform (row 3) were expressed from UAS inserts by elav^{C155}GAL4, and compared with wild type (raw 4). Third instar salivary glands were stained for speckles with anti-SC35 antibodies and dTET with anti-HA antibodies. The overlay of dTET and SC35 are shown in the merge. The boxplot on the right shows the quantification by image analysis of dTET accumulation in speckles. Values exceeding 1 (dashed black line) denote colocalization with SC35. Scale bar: 15 µm. Error bars represent the standard deviation.

isoform without this motif resembles vertebrate TET2 [87]. We have shown that human and mouse TET proteins localize to speckles (Figure 3A-C, Fig. S4-S8), which are sites for pre-mRNA splicing [88]. Next, we investigated TET subnuclear localization in Drosophila to clarify whether it is DNA methylation-dependent. For this purpose, we used Drosophila salivary glands expressing different TET isoforms (dTET short and dTET long), including a catalytically dead short isoform (dTET shortCD), using C-terminally HA-tagged UAS (Upstream Activation Sequence) constructs expressed with elavC155GAL4. In the large Drosophila salivary gland cells, speckles can also be visualized with antibodies against human SRSF2 (SC35) [89,90]. Interestingly, by performing immunofluorescence analysis we found that all three dTETs colocalize with speckles in the nucleus but were not detected in the nucleolus and that this localization is not dependent on the catalytic activity of dTET (Figure 3D).

Then, to analyze how the loss of TET affects gene expression in Drosophila, we analyzed a differential gene expression RNA-seq data set for changes, finding 594 and 326 two-fold up- or down-regulated genes, respectively (Supplementary Data S3) [67]. This analysis further revealed significant changes in alternative splicing (> 30% changes and $p < 0.01$) on exon skipping in 27 genes (Figure 4A), intron retention in 16 genes (Figure 4B), of 5' and 3' splice sites for 42 and 19 genes, respectively (Figure 4C-D), and 26 mutually exclusively spliced exons in 7 genes (Supplementary Data S4). Among changes in the genes with mutually exclusive spliced exons, we found several events for the Dscam1 gene. In Drosophila, this is one of the most complex alternatively spliced genes harboring three variable clusters, exon 4, exon 6, and exon 9, with 12, 48, and 33 variables, respectively [91]. Furthermore, gene ontology (GO) analysis of differentially expressed genes revealed a significant number of genes involved in detoxification including several Cytochrome C genes and metabolic genes (Figure 4E). GO analysis for changes focused on alternative 5' splice sites revealed a significant enrichment ($p < 0.05$) in neurodevelopmental genes, which is consistent with the defects in TETnull mutants [41,92].

Splicing speckles localize in the proximity of transcription sites and form subcellular domains called gene expression factories [88]. Since Drosophila lacks 5mC in DNA [93] but still has a TET gene highly homologous to the vertebrate TET proteins, we hypothesized a DNA methylation-independent function of TET proteins in chromatin-associated regulation of transcription. Interestingly, we found a colocalization of TET with elongating RNA polymerase II (RNAPII) in Drosophila salivary glands (Fig. S9A and Fig. S9B). TET localization at splicing speckles and TET colocalization with RNAPII were found in all conditions in Drosophila cells. The same colocalization between TET1 and RNAPII was found in mouse J1 mESCs, in addition to TET1 localization at splicing speckles (Fig. S9C-D). These results show that both elongating RNAPII and TET colocalize, with TET located at splicing speckles in the so-called 'gene expression factories.' Therefore, TET proteins may have a conserved role in regulating mRNA splicing and gene expression by alternative splicing. Interestingly, the RNAPII CTD has been shown to coordinate pre-mRNA processing events [94].

TET1 association with speckles and splicing factors interaction is RNA-dependent

To study the characteristics of the association of TET1 with splicing speckles and whether it is dependent on the interaction with splicing factors or with RNA, we performed live-cell RNase extraction followed by immunostaining in mouse tail fibroblast (MTF) cells. Representative images obtained by confocal microscopy are shown in Figure 5A, illustrating the total reduction of TET1 signal in extracted cells treated with RNase. In addition, the quantification of propidium iodide staining (which stains both DNA and RNA) shows a clear reduction after RNase treatment and extraction, corresponding to the RNA being washed out (Figure 5A-right). Therefore, the association of TET1 with splicing speckles is completely lost after RNA removal, while speckles are still visible based on the SC35 signal. This suggests that TET1's association with splicing speckles is primarily RNA-dependent. Therefore, RNA presence is a prerequisite for the reported colocalization and protein–protein interaction with splicing factors like U2AF.

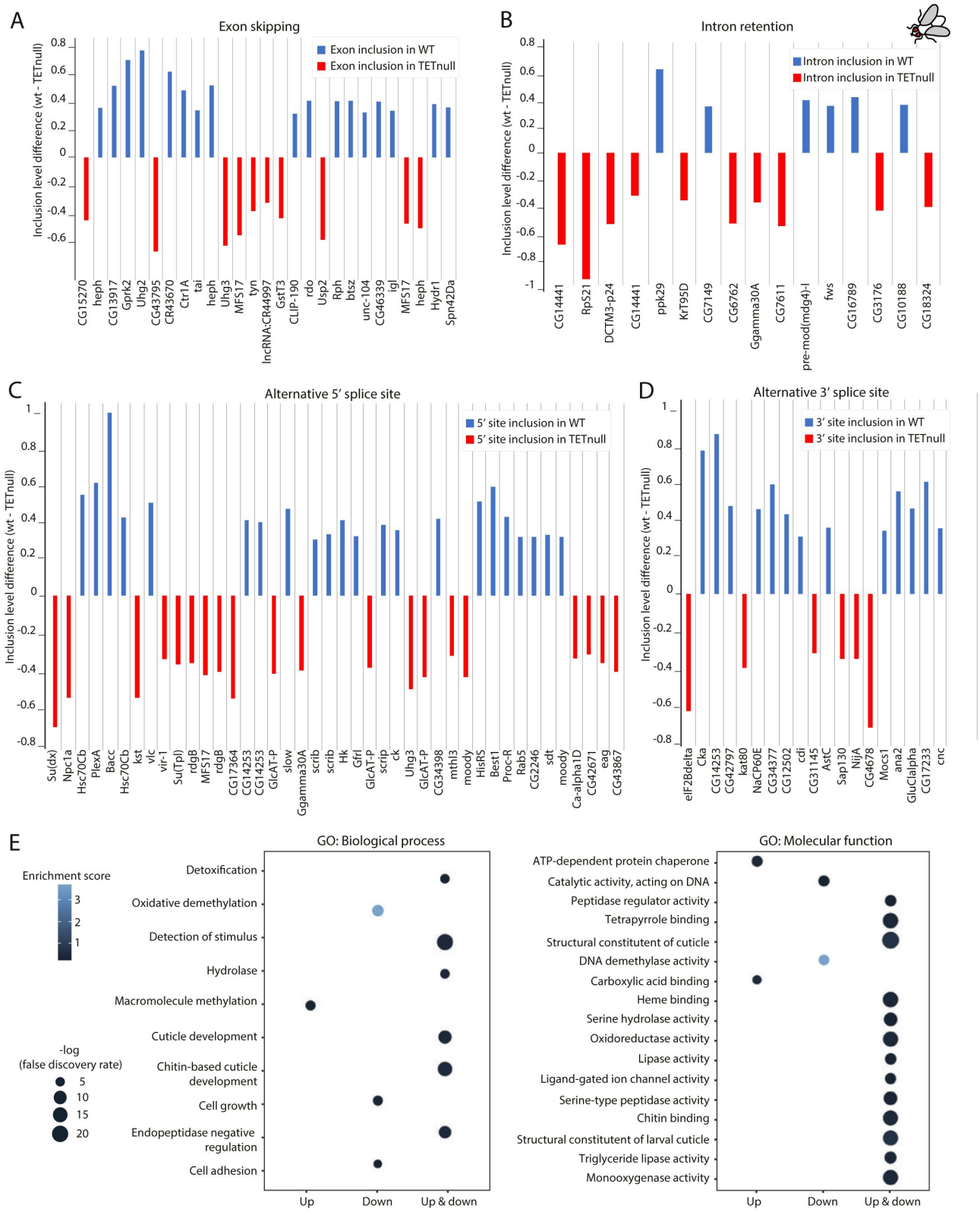


Figure 4. RNA-seq analysis of alternative splicing in Drosophila wild-type versus TETnull. (A) barplot showing the inclusion level difference of exon skipping genes between wild-type (wt) and TETnull samples. (B) barplot showing the inclusion level difference of intron retention between wild-type (wt) and TETnull samples. (C) barplot showing the inclusion level difference of alternative 5′ splice sites between wild-type (wt) and TETnull samples. (D) barplot showing the inclusion level difference of alternative 3′ splice sites between wild-type (wt) and TETnull samples. For (A), (B), (C), and (D), genes showing significant changes are shown (> 30% changes and $p < 0.01$, Supplementary data S4), ordered from lowest to highest p-value from left to right. The inclusion level

To verify this observation, we performed a fluorescent three-hybrid assay [70] in live cells to test the direct interaction of TET1-CD with an mRNA mimic. In this essay, the mRNA mimic (pms2-PABPC1-mCherry) is cotransfected in BHK cells [95] together with an RNA trap (MCP-EGFP-LacI), and the protein of interest, in this case PABPC1-mCherry (known to interact with the mRNA mimic) [70] or mCherry-TET1-CD. The RNA trap is tethered to the LacO locus in this cell line, which is visible as a big focus due to the EGFP fluorescence. This mRNA mimic is trapped at the LacO locus, and in the case of RNA–protein interaction, the recruitment of the target proteins by the mRNA mimic takes place. When this occurs, mCherry accumulation at the LacO can be imaged and quantified by live cell confocal microscopy (Figure 5B). After image analysis, we found a significant increase in the relative accumulation of TET1-CD in the LacO, as a result of its interaction with the mRNA mimic, compared to the negative control (without the mRNA and showing no mCherry accumulation at all), and with the positive control also with PABPC1-mCherry accumulation (Figure 5C-D). Thus, TET1 can be recruited by an mRNA mimic in live cells, confirming that TET1's association with speckles is RNA-driven.

In addition, we investigated the effect of RNA removal in the AlphaFold structural models, selecting the highest-scored models from Figure 2. As shown in the heatmaps of Figure 5E, almost all ipTMs and pTMs scores of previous predictions dropped to much lower values after removal of the short RNA sequence, demonstrating that RNA presence plays a crucial role in the interaction between TET and splicing factors (Supplementary Data S2). Structural modeling also suggests that the TET1 catalytic domain alone is predicted to interact with RNA, and this interaction displays similar confidence scores to AlphaFold3 structural predictions for TET1-DNA interaction (Figure 5F and Fig. S10). However, whether and how TET-RNA interaction affects RNA secondary structure needs to be further investigated.

Next, we investigated whether TET1 proteins affect hm5C levels in J1 mESCs, wild-type, and different TET1 deletion mutants (Fig. S11A). Previous studies have shown that hm5C levels in embryonic stem cell RNA depend on the presence of TET proteins, decreasing significantly in TET null mice [37]. Therefore, we quantified hm5C levels in total RNA extracted from wild-type mESCs by slot blot analysis. We performed these experiments by comparing untreated RNA samples with samples boiled at 65°C, RNA samples treated with RNase A, and plasmid DNA as negative controls. Interestingly, quantification of hm5C levels showed a two-fold increase for samples boiled at 65°C compared with unboiled samples (Fig. S11B). This suggests that hm5C in RNA is likely found in temperature-sensitive secondary structures. Similar results were published for f5C [96], which was shown to favor RNA duplex structures. In addition, we investigated whether different structural domains of TET1 affect the abundance of hm5C in RNA. In addition, we performed slot-blot using total RNA isolated from wild-type mESCs and mESCs expressing different TET1 deletion mutants (Supplementary Table S2). The quantification showed no major changes in hm5C levels in the total RNA of these cell lines. RNA hydroxymethylation levels were not negatively affected by TET1 N-terminal deletions in these cell lines (Fig. S11C), which still share a catalytic domain. In agreement with our previous results, this domain alone is sufficient for RNA binding and m5C oxidation. In addition, all TET1 N-terminal deletion mutants showed a clear colocalization of TET1 with SC35 (Fig. S11D). The studies performed in TET1 deletion mutant cells corroborated AlphaFold structural models, which identified a region in the CRD domain of TET

measures the frequency of a particular event (exon skipping, intron retention, 5′ or 3′ alternative splicing sites) in the final mRNA transcript, represented as a value between 0 (always skipped) and 1 (always included). The inclusion level difference is calculated as the inclusion level in wild-type minus the inclusion level in TETnull condition. Blue bars with positive values indicate an increased inclusion in the wild type, while red bars with negative values indicate an increased inclusion in TETnull. (E) gene ontology (GO) analysis of differentially expressed genes (Supplementary data S3) is shown on the left for biological process and on the right for molecular function, with enrichment indicated in blue color and -log(false discovery rate) by circle size.

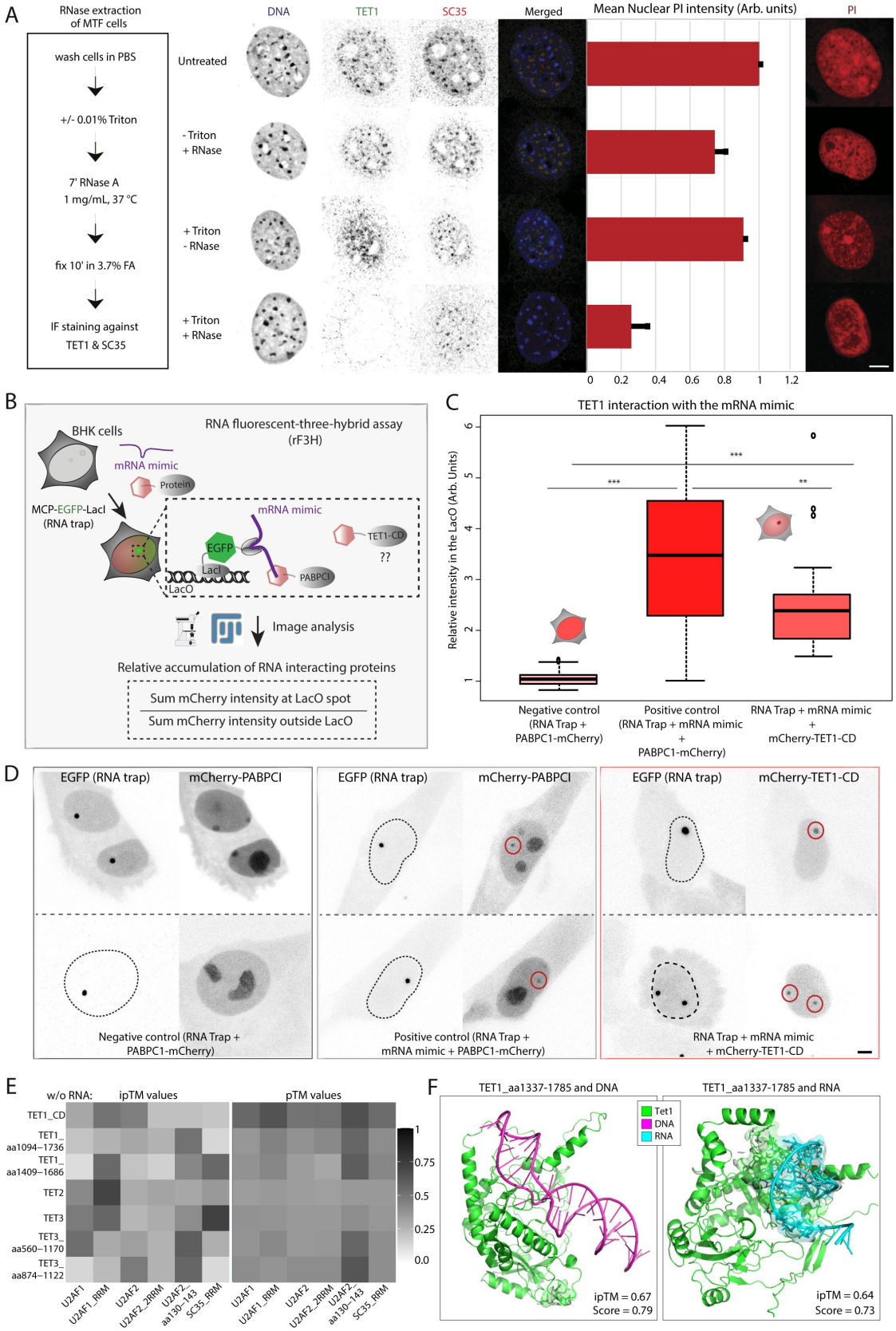


Figure 5. TET1 associates with RNA in splicing speckles and is recruited by an mRNA mimic in vivo. (A) live cell RNase extraction of MTF cells. On the left, the scheme shows the experimental pipeline to study TET1 speckle association upon RNase a treatment. Mouse tail fibroblasts were seeded on gelatine-treated glass coverslips and 24 hours later pre-extracted with 0.01% Triton in PBS, followed by incubation with 1 mg/mL RNase a for seven minutes. Afterward, cells were fixed with paraformaldehyde and

proteins to be involved in their interaction with U2AF. Altogether, this set of experimental approaches verified the interaction between TET1 catalytic domain (TET1-CD) and RNA, demonstrating the role of this protein in RNA metabolism and modifications.

TET proteins and m5C to hm5C oxidation promote splicing in vitro

Finally, we investigated whether TET proteins and/or their catalytic activity can affect RNA splicing. For this purpose, we used the nonsense-mediated decay (NMD) reporter for alternative splicing in mESCs wild type and knockouts for all three TET proteins or Dnmt. Alternative pre-mRNA splicing is a fundamental regulatory process for most mammalian multi-exon genes to increase proteome diversity. Nonsense-mediated mRNA decay (NMD) is a conserved mRNA surveillance mechanism to mitigate deleterious effects caused by gene mutations or transcriptional errors (reviewed in [97]. This pathway is initiated by the removal of both the poly(A) tail and the 5′ cap. The latter was shown to be coupled with translational repression [98]. NMD affects mRNAs containing premature translation termination codons (PTCs), which recruit the ATP-dependent RNA helicases Upf1–34. Based on the NMD system and using a transitory transfected NMD reporter vector for EGFP, if alternative splicing takes place, the mRNA encoding EGFP will be translated, and fluorescence intensity in the cells can be measured. If alternative splicing does not occur, the resulting non-sense mRNA will be degraded and EGFP will not be expressed (Figure 6A). We transfected different cell lines (wild type, TET triple knockout, and DNMT triple knockout) with a vector encoding the NMD reporter and different cherry-tagged TET constructs, including TET1-CDm (catalytically dead mutant), TET1-CD, TET2-CD, and TET3-CD. TET1-CDm harbors a mutation in the Fe (II)-binding motif and maintains its DNA binding ability but is catalytically inactive. We used the catalytic domain of TET proteins based on our previous protein-protein-RNA interaction results (Figures 2, 5, and Fig. S11). In addition, this domain is sufficient to oxidize m5C in single-stranded RNA [37]. After ectopic expression of the vectors, we found an increase in EGFP intensity for TET triple KO cell lines expressing TET1, TET2, and TET3 proteins, including TET1 catalytic dead mutant. Higher EGFP fluorescence relates to higher splicing levels for TET-transfected cell lines compared with controls transfected with mCherry only (Figure 6B).

We further investigated whether the effect of TET proteins promoting splicing was affected by DNA/RNA 5-methylcytosine levels (5mC/m5C). Both 5mC and m5C can be the substrate of TET proteins, necessary for 5mC/m5C to 5hmC/hm5C conversion. On the other hand, DNMT proteins are responsible for 5mC production, and it has been shown that DNMT2 has double-substrate specificity and adopts a similar catalytic mechanism to methylate RNA [99]. Therefore, we used mESCs triple knockouts of DNMT proteins, with residual levels of 5-methylcytosine [100,101]. We investigated whether the transitory transfection of TET proteins (as in Figure 6B) positively affected splicing levels in the absence of

immunostained with antibodies against TET1 and SC35. Representative confocal microscopy images are shown in the center. As treatment control, similarly treated cells were stained with propidium iodide (PI), and the respective signal was quantified by high-content microscopy ($n > 1000$ cells). Representative images of PI staining are shown on the right. Scale bar = 5 µm. (B) TET1-CD is recruited with a mRNA mimic in live-cell rF3H assay (RNA fluorescent three-hybrid assay). Scheme of the rF3H assay. Briefly, genetically modified BHK cells with many copies of the LacO were transfected with an RNA trap, an mRNA mimic, and pPABPCI-mCherry or pmCherry-TET1-CD. The RNA trap is fused to LacI (for LacO targeting) and EGFP (for visualization of the LacO locus). After live-cell imaging of transfected cells, the interaction between the mRNA mimic and the mCherry proteins is quantified by image analysis. (C) boxplot showing the quantification of the relative accumulation of mCherry proteins at the LacO. Two independent replicates were performed. N-number = 26–37. Statistical significance was tested with a paired two-sample wilcoxon test (n.S., not significant, is given for p-values ≥0.05; one star (*) for p-values < 0.05 and ≥ 0.005; two stars (**) is given for values < 0.005 and ≥ 0.0005; three stars (***) is given for values < 0.0005). P-values are shown in Supplementary Table S5. (D) Representative images of the rF3H assay for each condition. The area of the LacO locus is highlighted in the mCherry channel where the accumulation of mCherry proteins in the LacO is visible. Scale bar = 5 µm. (E) heatmaps showing ipTms and pTms scores of AlphaFold protein–protein interaction predictions between TET proteins and splicing factors without RNA. Higher-ranked predictions shown in Fig. 2 were repeated without the short RNA, to show RNA-specific effects on the prediction score. (F) structure showing AlphaFold 3 prediction model for TET1-CD and DNA (LINE1 5′ UTR region) versus TET1 interaction with RNA (intron-exon transition region).

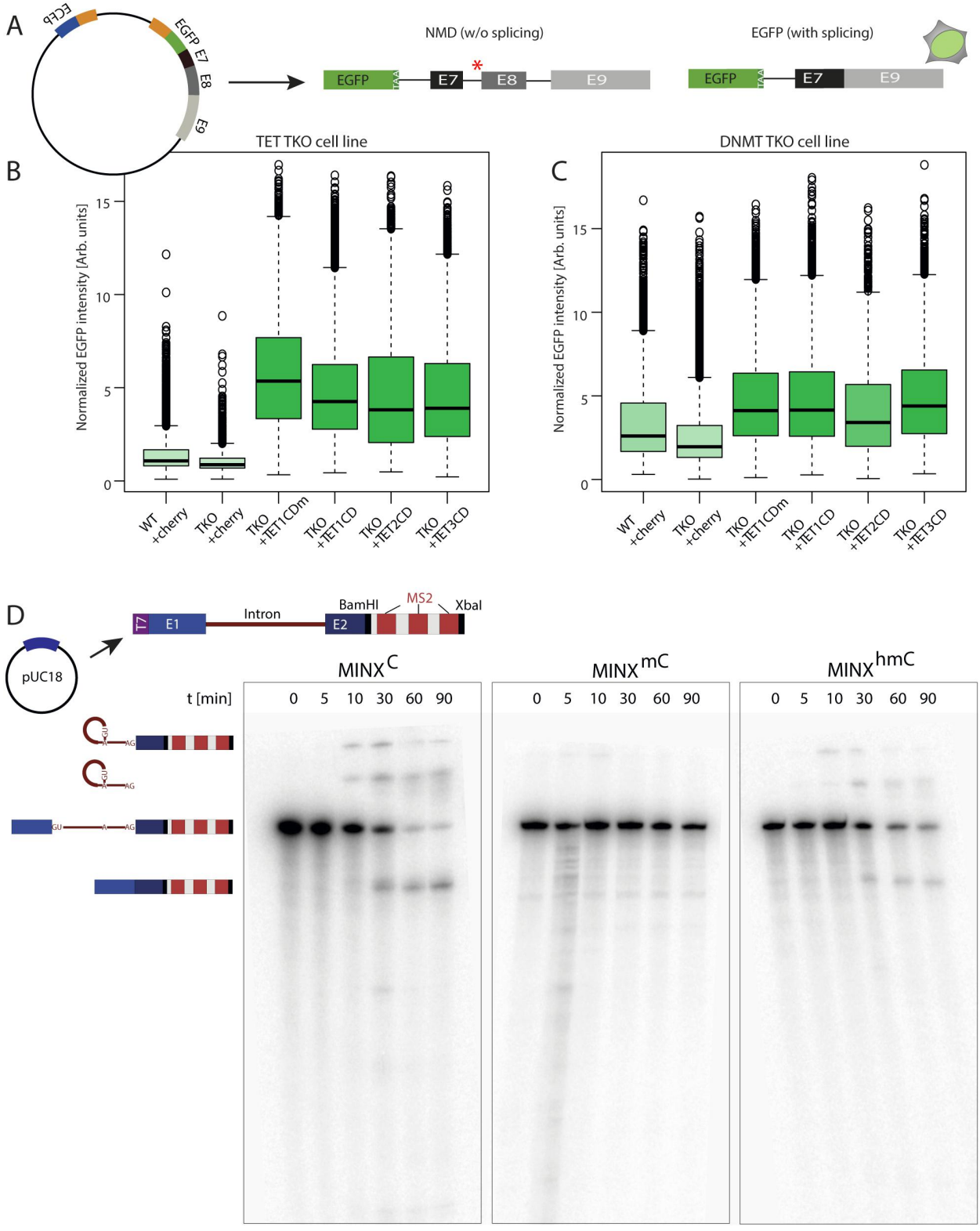


Figure 6. TET proteins and m5C to hm5C oxidation promote splicing. (A) scheme showing the chimeric vector used for the splicing assay based on the NMD system (non-sense mediated mRNA decay). E7, E8, and E9 correspond to exons 7, 8, and 9 of the HNRNPDL gene (heterogeneous nuclear ribonucleoprotein D-like), represented by black/gray rectangles. Introns between the exons are represented by a black line. The red star indicates premature translation termination codons (PTCs) present in the chimeric 'pre-mRNA.' If splicing takes place, the PTC is removed, and translation of the RNA, with GFP at the 5,' occurs. In the absence of splicing, the PTC leads to the degradation of the 'pre-mRNA.' Stop codon 'TAA' at the end of the GFP coding sequence. (B) boxplots showing the normalized EGFP intensity in different mEscs (wild type and TET triple knockouts) transfected with the NMD vector and the

5-methylcytosine. Following the trend of the previous experiment, we also found an increase in the EGFP intensity compared with controls (only mCherry transfected cells) corresponding with an increase in splicing in the presence of TET proteins. In this case, the fold increase was slightly lower than in Figure 6B using TET triple knockout cell lines (Figure 6C). This can be explained by higher levels of endogenous TET proteins in these cell lines, already affecting the splicing of the NMD reporter to certain levels, or due to the effect of TET-mediated m5C oxidation favoring splicing in addition to TET presence alone. Mouse ESCs transfected with TET1-CD or TET1-CDm (catalytically dead mutant) showed similar intensity values for EGFP, meaning increased splicing for both (Figure 6B–C). Therefore, the first option seems more likely. Nonetheless, the levels of alternative splicing still increase for all TET-transfected cells independently of the absence of 5-methylcytosine (in triple DNMT KO cells) or TET1 catalytic activity (TET1-CDmutant). These results show that TET proteins can promote splicing independently of their catalytic activity.

Alternatively, we studied splicing levels using a different splicing reporter system, the spliceable firefly luciferase with one or two introns (human hemoglobin alpha [hHBA] and human hemoglobin beta [hHBB]) under the control of a CMV promoter. Using this system, the relation between the number of introns and splicing performance can be studied in more detail (Fig. S12A). HEK cells were co-transfected with firefly luciferase and mCherry, mCherry-TET1-CD, or mCherry-TET1-CDm. To quantify splicing events, extracts of cells ectopically expressing the desired constructs were mixed with a substrate metabolizable by luciferase. When the mRNA is spliced correctly, the enzymatic activity of luciferase results in a luminescent signal. Then, this luminescence was measured and normalized to the fluorescent signal of mCherry. In these experiments, the luminescence signal was low irrespective of the number of introns for control cells (transfected only with mCherry). However, there was an increase in luminescence signal in cells transfected with both TET1-CD and TET1-CDm, corresponding with an increase in splicing efficiency (Fig. S12B-C). The luminescence signal was much higher when the mRNA comprised a single intron (green bars), compared with the mRNA with two introns (blue bars), indicating that splicing efficiency is almost threefold higher when only one intron has to be removed. In this case, splicing efficiency is slightly favored by TET1 catalytic activity compared to TET1-CDm. These results suggest that the binding of TET1 alone influences the splicing efficiency of luciferase mRNA, independently of the catalytic activity of TET1. In summary, TET proteins interact with U2AF promoting splicing, an activity that is neither dependent on their catalytic activity nor the substrate (m5C) (Figure 6 and Fig. S12). However, TET catalytic activity and oxidation of m5C may still play a specific role in favoring splicing efficiency.

To further investigate this, we studied the effect of cytosine modification levels (m5C and hm5C) in RNA splicing. The luciferase splicing reporter showed a slightly higher increase in splicing efficiency for TET1-CD compared to TET1-CDm (Fig. S12B). Therefore, we investigated whether RNA methylation levels and m5C oxidation specifically affect splicing efficiency. For this purpose, we performed an in vitro splicing assay using MINX-M3 pre-mRNA reporter construct [73]. This vector consists of a pre-mRNA encoding adenoviral major late gene sequences exons 1 and 2 interrupted by an intron and tagged with 2xMS2 bacteriophage coat protein (Figure 6D). For this purpose,

different TET constructs. The same experiments were performed in Dnmt triple knockout cell lines, and quantification is shown in (C). (D) in vitro splicing assay. On the top is the scheme of the vector used for in vitro splicing. The pre-mRNA encoding adenoviral major late gene sequence exons 1 and 2 interrupted by an intron was created by T7 RNA polymerase run-off transcription. MS2 bacteriophage coat protein binding aptamer (not relevant for this study). For this reaction, the only cytidine sources were either CTP, m5CTP, or hm5CTP. 32p-labeled transcripts were phenol-chloroform purified, and similar RNA amounts were subjected to in vitro splicing reactions with HeLa nuclear extracts (0, 5, 10, 30, 60, 90 min). A diagram representing the different splicing intermediates of this construct is shown on the bottom left. Samples were taken at the indicated time points and analyzed on polyacrylamide gels.

MINX-M3 pre-mRNA was produced with run-off IVT (in vitro transcription) reactions containing either CTP, m5CTP, or hm5CTP, as the only cytidine. 32P-labeled transcripts were subjected to in vitro splicing using HeLa nuclear extracts, a tumor-derived cell line with high levels of TET1 protein [102]. The time course of the splicing reaction was analyzed on polyacrylamide gels (Figure 6D). Similar splicing kinetics and efficiency were observed for the unmodified and hm5C-containing pre-mRNAs, while no splicing was detected when pre-mRNA containing m5C was used. These results indicate that cytosine methylation affects splicing negatively in vitro, but this can be restored by oxidation of mC to hmC. Thus, RNA modifications like m5C and hm5C also may play a role in regulating splicing efficiency. In a previous publication, DNA base modifications have been shown to affect DNA physical properties and DNA processes, with cytosine methylation stabilizing the DNA helix and increasing its melting temperature. The latter affects DNA helicases and RNA/DNA polymerase speed. In this scenario, the oxidation of methylated cytosines reverses the duplex stability and genome metabolic effects of the unmodified cytosine [103]. The latter may be extrapolated to RNA modifications also playing a structural role in RNA metabolism.

Together with our finding that hm5C is predominantly found in temperature-sensitive secondary structures, we speculate on the biological role of RNA hydroxymethylation in topologically regulating splicing events. In this way, splicing would be more likely to occur depending on the methylation status of the pre-mRNA, which may form different secondary structures depending on its physical properties. The latter would be related to the variegated deposition of certain cytosine modifications along the pre-mRNA sequence which in turn may rely on TET activity.

Conclusions and outlook

This study highlights the multifaceted and undisclosed role of TET dioxygenase proteins in RNA metabolism, particularly in regulating RNA splicing in mammalian and Drosophila cells. We demonstrate that TET proteins, beyond their well-established role in DNA demethylation, localize to splicing speckles and interact with key splicing factors such as U2AF1 and U2AF2. These interactions enhance splicing efficiency, independent of TET catalytic activity, as we have shown using different splicing reporter systems and RNA-seq analysis. In this regard, there is previous evidence showing that non-catalytic functions of TET1 are critical for the regulation of gene expression and the silencing of endogenous retroviruses, being an interaction hub for multiple chromatin-modifying complexes [104]. However, we also show that splicing efficiency in vitro is reduced in the presence of m5C, while hm5C is able to restore splicing efficiency. Based on the enrichment of TET to splicing speckles in vivo, as well as the outcome of the cellular splicing assays, we propose that this oxidation may be mediated by TET proteins. This suggests a dual role for TET proteins in splicing: a structural or scaffolding function and a catalytic, RNA modification-dependent function. This dual role may be interconnected, having implications for understanding post-transcriptional and alternative splicing regulation in various biological contexts, including development and disease. Interestingly, it has been shown that a big fraction of RNA splicing regulators are involved in neurogenesis, playing a role in neural development [15]. In this process, the precise regulation of alternative splicing is crucial [105]. TET seems to arise as one of these regulators, as the RNA-seq analysis of splicing events in Drosophila shows. The enrichment of changes on alternative 5' splice sites for neurodevelopmental genes and the defects in TETnull mutants support this hypothesis.

The distribution of m5C varies among cell types [106], and modifications at specific positions on mRNA can have different regulatory roles promoting or inhibiting translation [107,108]. As we have shown, TET1 association with splicing speckles is RNA dependent, not existing a direct interaction with the splicing factor SC35 (SRSF2). The latter could be interpreted as possible competition of TET and SC35 for binding sites since this protein has been reported to read m5C in mRNA [109]. In addition, TET was shown to oxidize m5C in polyadenylated RNAs in Drosophila, favoring mRNA

translation. TET and RNA hydroxymethylation were found to be most abundant in the Drosophila brain, with TET-deficient flies suffering from impaired brain development [41]. The finding that TET proteins can oxidize 5mC in RNA to promote translation in Drosophila, together with our results in mouse, human, and Drosophila, point to a conserved role of TET proteins in RNA metabolism, affecting splicing efficiency and therefore subsequent translation of transcripts. In this regard, TET proteins arise as a modifier of m5C, in addition to their canonical role in DNA 5mC oxidation. Future research should explore the mechanistic details of TET's interaction with the spliceosome and the broader impact of RNA hydroxymethylation on RNA structure and function.

In the past, the link between TET1 and the NuRD chromatin remodeling complex was reported [110], which mediates gene repression to maintain the pluripotent state of ESCs. In addition, we have shown the subnuclear localization of the short isoform of TET1, lacking the zinc finger domain, to heterochromatic regions during the late S phase [62]. These findings suggest that TET1 switches between chromatin-associated and RNA-associated localizations, the latter independent of the cell cycle state. Furthermore, it has been shown that the presence of hm5C in the transcribed gene promotes the annealing of the nascent RNA to the template DNA strand, leading to the formation of an R-loop. The resolution of these R-loops leads to differential expression of a subset of genes involved in stem cell proliferation. In this context, TET activity promotes co-transcriptional R-loop formation as a mechanism of gene expression regulation [111]. The latter may be particularly related to our finding that hm5C in RNA seems to be located in secondary structures. Strikingly, TET localization at speckles was found across different cell lines, independent of the differentiation state and cell fate. It is also conserved between TET proteins and between mouse, human, and Drosophila cells, the latter without DNA methylation. In this scenario, it is tempting to speculate that the function of TET in splicing reflects an ancestral and perhaps primary role in RNA metabolism.

Highlights

- TET1 localizes to splicing speckles in an RNA-dependent manner
- TET proteins, especially TET1, interact with the splicing factors U2AF1 and U2AF2
- TET proteins increase splicing efficiency, independent of their catalytic activity
- RNA 5-methylcytosine (m5C) oxidation to 5-hydroxymethylcytosine (hm5C) restores splicing efficiently in vitro

Acknowledgments

We thank the bioinformatics core facility of the Institute of Molecular Biology (IMB) for providing support in AlphaFold 2 modeling. We thank Reinhard Lührmann for his advice and help with the in vitro splicing assay. We also thank Florian Heyd for the U2AF1 and 2 plasmids. We are indebted to Beatrix Suess, and Marc Vogel for kindly providing materials for the cellular splicing assay. We thank BacPac for BAC clones, L. Waltzer for TET mutants, the Bloomington stock center for flies, and FlyORF for injections.

F.D.H., J.W., C.B., M.A., A.Z., D.N.D.S., T.C.D, and R.A. performed experiments. F.D.H., J.W., C.B., M.A., A.Z., D.N.D.S., T.C.D, and R.A. analyzed data. F.D.H. developed data analysis pipelines. M.A. performed structural modeling with AF and data analysis. S.B., H.L., and M.S. provided tools and advice. F.D.H., M.A., and M.C.C. conceived and developed the project. M.A. generated the final figures and wrote the manuscript. All authors agreed on the manuscript and contributed to the editing of the manuscript.

Author contributions

CRediT: **Florian D. Hastert:** Conceptualization, Investigation, Methodology, Writing – review & editing; **Jasmin Weber:** Investigation, Writing – review & editing; **Christina Bauer:** Investigation, Writing – review & editing; **Andreas Zhadan:** Investigation; **Deepanshu N. D. Singh:** Investigation; **Thomas C. Dix:** Investigation; **Roland Arnold:** Investigation; **Sergey Bessonov:** Investigation, Writing – review & editing; **Matthias Soller:** Funding acquisition, Writing – review & editing; **Heinrich Leonhardt:** Writing – review & editing; **M. Cristina Cardoso:** Conceptualization, Funding acquisition, Resources, Writing – review & editing; **Maria Arroyo:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 393547839 – SFB 1361, CA 198/16–1 project number 425470807 and CA 198/19–1 project number 522122731 to M.C.C. and by the Biotechnology and Biological Science Research Council to M.S.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

All data, including unprocessed images and source datasets, have been deposited and are publicly available at TuDatalib (https://doi.org/10.48328/tudatalib-1657.3). Biological materials will be made available upon request from the corresponding author Maria Arroyo (arroyo.lopez.mc@gmail.com) and M. Cristina Cardoso (cardoso@bio.tu-darmstadt.de). Plasmid constructs are also available via Addgene (see Supplementary Table S1). Mass spectrometry data can be found in Supplementary Data S1. Information about AlphaFold3 scores can be found in Supplementary Data S2. RNA-seq analysis processed files can be found in Supplementary Data S3 (differential gene expression) and Supplementary Data S4 (splicing events). Curated and annotated unprocessed images for Western blot are available in Supplementary Data S5. Published RNA-seq data sets used in this study can be found in GEO (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi) under the accession number GSE206852.

ORCID

Florian D. Hastert http://orcid.org/0000-0001-9961-5766
Maria Arroyo http://orcid.org/0000-0002-5837-8060

References

[1] Thomas J, Palusa SG, Prasad KVSK, et al. Identification of an intronic splicing regulatory element involved in auto-regulation of alternative splicing of SCL33 pre-mRNA. The Plant Journal. 2012;72 (6):935–946. doi: 10.1111/tpj.12004

[2] Wansink DG, Schul W, van der Kraan I, et al. Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. The Journal Of Cell Biology. 1993;122(2):283–293. doi: 10.1083/jcb.122.2.283

[3] Galganski L, Urbanek MO, Krzyzosiak WJ. Nuclear speckles: molecular organization, biological function and role in disease. Nucleic Acids Research. 2017;45 (18):10350–10368. doi: 10.1093/nar/gkx759

[4] Spector DL, Fu XD, Maniatis T. Associations between distinct pre-mRNA splicing components and the cell nucleus. Embo J. 1991;10(11):3467–3481. doi: 10.1002/j.1460-2075.1991.tb04911.x

[5] Spector DL, Schrier WH, Busch H. Immunoelectron microscopic localization of snRnps. Biol Cell. 1983;49 (1):1–10. doi: 10.1111/j.1768-322X.1984.tb00215.x

[6] Perraud M, Gioud M, Monier JC. Intranuclear structures of monkey kidney cells recognised by immunofluorescence and immuno-electron microscopy using anti-ribonucleoprotein antibodies (author's transl). Ann Immunol (Paris). 1979;130C (5):635–647.

[7] Zamore PD, Patton JG, Green MR. Cloning and domain structure of the mammalian splicing factor U2AF. Nature. 1992;355(6361):609–614. doi: 10.1038/355609a0

[8] Carmo-Fonseca M, Pepperkok R, Carvalho MT, et al. Transcription-dependent colocalization of the U1, U2, U4/U6, and U5 snRnps in coiled bodies. J Cell Biol. 1992;117(1):1–14. doi: 10.1083/jcb.117.1.1

[9] Fu XD. The superfamily of arginine/serine-rich splicing factors. RNA. 1995;1(7):663–680.

[10] Wu JY, Maniatis T. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell. 1993;75(6):1061–1070. doi: 10.1016/0092-8674(93)90316-I

[11] Kashyap A, Tripathi G, Tripathi A, et al. RNA splicing: a dual-edged sword for hepatocellular carcinoma. Med Oncol. 2022;39(11):173. doi: 10.1007/s12032-022-01726-8

[12] Wang E, Aifantis I. RNA splicing and cancer. Trends Cancer. 2020;6(8):631–644. doi: 10.1016/j.trecan.2020.04.011

[13] Bradley RK, Anczuków O. RNA splicing dysregulation and the hallmarks of cancer. Nat Rev Cancer. 2023;23 (3):135–155. doi: 10.1038/s41568-022-00541-7

[14] Conboy JG. RNA splicing during terminal erythropoiesis. Curr Opin Hematol. 2017;24 (3):215–221. doi: 10.1097/MOH.0000000000000329

[15] Fisher E, Feng J. RNA splicing regulators play critical roles in neurogenesis. Wiley Interdiscip Rev RNA. 2022;13(6):e1728. doi: 10.1002/wrna.1728

[16] Martin KC, Ephrussi A. mRNA localization: gene expression in the spatial dimension. Cell. 2009;136 (4):719–730. doi: 10.1016/j.cell.2009.01.044

[17] Karikó K, Muramatsu H, Welsh FA, et al. Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. Mol Ther. 2008;16(11):1833–1840. doi: 10.1038/mt.2008.200

[18] Mendel M, Delaney K, Pandey RR, et al. Splice site m6A methylation prevents binding of U2AF35 to inhibit RNA splicing. Cell. 2021;184(12):3125–3142.e25. doi: 10.1016/j.cell.2021.03.062

[19] Shenasa H, Bentley DL. Pre-mRNA splicing and its cotranscriptional connections. Trends Genet. 2023;39 (9):672–685. doi: 10.1016/j.tig.2023.04.008

[20] Trojer P, Reinberg D. Facultative heterochromatin: is there a distinctive molecular signature? Mol Cell. 2007;28(1):1–13. doi: 10.1016/j.molcel.2007.09.011
[21] Bestor TH. The DNA methyltransferases of mammals. Hum Mol Genet. 2000;9(16):2395–2402. doi: 10.1093/hmg/9.16.2395
[22] Bird A. DNA methylation patterns and epigenetic memory. Genes Dev. 2002;16(1):6–21. doi: 10.1101/gad.947102
[23] Rácz I, Király I, Lásztily D. Effect of light on the nucleotide composition of rRNA of wheat seedlings. Planta. 1978;142(3):263–267. doi: 10.1007/BF00385075
[24] Motorin Y, Lyko F, Helm M. 5-methylcytosine in RNA: detection, enzymatic formation and biological functions. Nucleic Acids Res. 2010;38(5):1415–1430. doi: 10.1093/nar/gkp1117
[25] Huber LC, Ulrich S, Leuenberger C, et al. Featured article: microRNA-125a in pulmonary hypertension: regulator of a proliferative phenotype of endothelial cells. Exp Biol Med (maywood). 2015;240 (12):1580–1589. doi: 10.1177/1535370215579018
[26] Guo G, Pan K, Fang S, et al. Advances in mRNA 5-methylcytosine modifications: detection, effectors, biological functions, and clinical relevance. Mol Ther Nucleic Acids. 2021;26:575–593. doi: 10.1016/j.omtn.2021.08.020
[27] Reid R, Greene PJ, Santi DV. Exposition of a family of RNA m(5)C methyltransferases from searching genomic and proteomic sequences. Nucleic Acids Res. 1999;27(15):3138–3145. doi: 10.1093/nar/27.15.3138
[28] Goll MG, Kirpekar F, Maggert KA, et al. Methylation of tRnaasp by the DNA methyltransferase homolog Dnmt2. Science. 2006;311(5759):395–398. doi: 10.1126/science.1120976
[29] Arzumanian VA, Dolgalev GV, Kurbatov IY, et al. Epitranscriptome: review of top 25 most-studied RNA modifications. Int J Mol Sci. 2022;23(22):13851. doi: 10.3390/ijms232213851
[30] Cappannini A, Ray A, Purta E, et al. MODOMICS: a database of RNA modifications and related information. 2023 update. Nucleic Acids Res. 2024;52(D1): D239–44. doi: 10.1093/nar/gkad1083
[31] Fu Y, Jia G, Pang X, et al. FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. Nat Commun. 2013;4(1):1798. doi: 10.1038/ncomms2822
[32] Jia G, Fu Y, Zhao X, et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol. 2011;7 (12):885–887. doi: 10.1038/nchembio.687
[33] Hu L, Li Z, Cheng J, et al. Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. Cell. 2013;155(7):1545–1555. doi: 10.1016/j.cell.2013.11.020
[34] Ito S, Shen L, Dai Q, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011;333(6047):1300–1303. doi: 10.1126/science.1210597
[35] Fedeles BI, Singh V, Delaney JC, et al. The AlkB family of Fe(II)/α-ketoglutarate-dependent dioxygenases: repairing nucleic acid alkylation damage and beyond. J Biol Chem. 2015;290(34):20734–20742. doi: 10.1074/jbc.R115.656462
[36] Falnes PØ, Bjørås M, Aas PA, et al. Substrate specificities of bacterial and human AlkB proteins. Nucleic Acids Res. 2004;32(11):3456–3461. doi: 10.1093/nar/gkh655
[37] Fu L, Guerrero CR, Zhong N, et al. Tet-mediated formation of 5-hydroxymethylcytosine in RNA. J Am Chem Soc. 2014;136(33):11582–11585. doi: 10.1021/ja505305z
[38] Huber SM, van Delft P, Mendil L, et al. Formation and abundance of 5-hydroxymethylcytosine in RNA. Chembiochem. 2015;16(5):752–755. doi: 10.1002/cbic.201500013
[39] Zhang H-Y, Xiong J, Qi B-L, et al. The existence of 5-hydroxymethylcytosine and 5-formylcytosine in both DNA and RNA in mammals. Chem Commun. 2016;52 (4):737–740. doi: 10.1039/C5CC07354E
[40] Kremer EA, Gaur N, Lee MA, et al. Interplay between TETs and microRNAs in the adult brain for memory formation. Sci Rep. 2018;8(1):1678. doi: 10.1038/s41598-018-19806-z
[41] Delatte B, Wang F, Ngoc LV, et al. RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. Science. 2016;351 (6270):282–285. doi: 10.1126/science.aac5253
[42] Li D, Wang Q, Bayat A, et al. Spliceosome malfunction causes neurodevelopmental disorders with overlapping features. J Clin Invest. 2024;134(1):e171235. doi: 10.1172/JCI171235.
[43] Vogel M, Weigand JE, Kluge B. A small, portable RNA device for the control of exon skipping in mammalian cells. Nucleic Acids Res. 2018;46(8):e48. doi: 10.1093/nar/gky062
[44] Peters AH, O'Carroll D, Scherthan H, et al. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell. 2001;107 (3):323–337. doi: 10.1016/S0092-8674(01)00542-6
[45] Ludwig AK, Zhang P, Hastert FD, et al. Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise. Nucleic Acids Res. 2017;45(5):2438–2457. doi: 10.1093/nar/gkw1197
[46] Stevens LC. A new inbred subline of mice (129-terSv) with a high incidence of spontaneous congenital testicular teratomas. J Natl Cancer Inst. 1973;50 (1):235–242. doi: 10.1093/jnci/50.1.235
[47] Dawlaty MM, Breiling A, Le T, et al. Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. Dev Cell. 2014;29(1):102–111. doi: 10.1016/j.devcel.2014.03.003
[48] Agarwal N, Becker A, Jost KL, et al. MeCP2 Rett mutations affect large scale chromatin organization. Hum Mol Genet. 2011;20(21):4187–4195. doi: 10.1093/hmg/ddr346

[49] Becker A, Zhang P, Allmann L, et al. Poly(ADP-ribosyl)ation of methyl CpG binding domain protein 2 regulates chromatin structure. J Biol Chem. 2016;291(10):4873–4881. doi: 10.1074/jbc.M115.698357
[50] Müller U, Bauer C, Siegl M, et al. TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. Nucleic Acids Res. 2014;42(13):8592–8604. doi: 10.1093/nar/gku552
[51] Bauer C, Göbel K, Nagaraj N, et al. Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosamine transferase (OGT). J Biol Chem. 2015;290(8):4801–4812. doi: 10.1074/jbc.M114.605881
[52] Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26(12):1367–1372. doi: 10.1038/nbt.1511
[53] Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630(8016):493–500. doi: 10.1038/s41586-024-07487-w
[54] Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. Proteins. 2004;57(4):702–710. doi: 10.1002/prot.20264
[55] Xu J, Zhang Y. How significant is a protein structure similarity with TM-score = 0.5? Bioinformatics. 2010;26(7):889–895. doi: 10.1093/bioinformatics/btq066
[56] Varadi M, Bertoni D, Magana P, et al. AlphaFold protein structure database in 2024: providing structure coverage for over 214 million protein sequences. Nucleic Acids Res. 2024;52(D1):D368–75. doi: 10.1093/nar/gkad1011
[57] Evans R, O'Neill M, Pritzel A, et al. Protein complex prediction with AlphaFold-Multimer. BioRxiv. 2021. doi:10.1101/2021.10.04.463034.
[58] Bateman A, Martin M-J, Orchard S, et al. Uniprot: the universal protein knowledgebase in 2023. Nucleic Acids Research. 2023;51(D1):D523–31. doi: 10.1093/nar/gkac1052
[59] Schrodinger. PyMOL the PyMOL molecular graphics system, Version 1.8. Schrödinger, LLC.
[60] Pettersen EF, Goddard TD, Huang CC, et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci. 2021;30(1):70–82. doi: 10.1002/pro.3943
[61] Cardoso MC, Joseph C, Rahn HP, et al. Mapping and use of a sequence that targets DNA ligase I to sites of DNA replication in vivo. J Cell Biol. 1997;139(3):579–587. doi: 10.1083/jcb.139.3.579
[62] Arroyo M, Hastert FD, Zhadan A, et al. Isoform-specific and ubiquitination dependent recruitment of Tet1 to replicating heterochromatin modulates methylcytosine oxidation. Nat Commun. 2022;13(1):5173. doi: 10.1038/s41467-022-32799-8
[63] Arroyo M, Cardoso CM, Hastert FD. In situ quantification of cytosine modification levels in heterochromatic domains of cultured mammalian cells. Bio Protoc. 2023;13(14):e4716. doi: 10.21769/BioProtoc.4716
[64] Haussmann IU, White K, Soller M. Erect wing regulates synaptic growth in Drosophila by integration of multiple signaling pathways. Genome Biol. 2008;9(4):R73. doi: 10.1186/gb-2008-9-4-r73
[65] Haussmann IU, Ustaoglu P, Brauer U, et al. Plasmid-based gap-repair recombineered transgenes reveal a central role for introns in mutually exclusive alternative splicing in down syndrome cell adhesion molecule exon 4. Nucleic Acids Res. 2019;47(3):1389–1403. doi: 10.1093/nar/gky1254
[66] Dix TC, Lassota A, Brauer U, et al. Gap-repair recombineering for efficient retrieval of large DNA fragments from BAC clones and manipulation of large high-copy number plasmids. BMC Methods. 2024;1(1). doi: 10.1186/s44330-024-00011-6
[67] Boulet M, Gilbert G, Renaud Y, et al. Adenine methylation is very scarce in the Drosophila genome and not erased by the ten-eleven translocation dioxygenase. Elife. 2023;12. doi: 10.7554/eLife.91655
[68] Haussmann IU, Wu Y, Nallasivan MP, et al. Cmtr cap-adjacent 2'-O-ribose mRNA methyltransferases are required for reward learning and mRNA localization to synapses. Nat Commun. 2022;13(1):1209. doi: 10.1038/s41467-022-28549-5
[69] Herce HD, Casas-Delucchi CS, Cardoso MC. New image colocalization coefficient for fluorescence microscopy to quantify (bio-)molecular interactions. J Microsc. 2013;249(3):184–194. doi: 10.1111/jmi.12008
[70] Duan N, Arroyo M, Deng W, et al. Visualization and characterization of RNA-protein interactions in living cells. Nucleic Acids Res. 2021;49(18):e107. doi: 10.1093/nar/gkab614
[71] Zillmann M, Zapp ML, Berget SM. Gel electrophoretic isolation of splicing complexes containing U1 small nuclear ribonucleoprotein particles. Mol Cell Biol. 1988;8(2):814–821. doi: 10.1128/MCB.8.2.814
[72] Zhang P, Hastert FD, Ludwig AK, et al. DNA base flipping analytical pipeline. Biol Methods Protoc. 2017;2(1):bpx010. doi: 10.1093/biomethods/bpx010
[73] Deckert J, Hartmuth K, Boehringer D, et al. Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. Mol Cell Biol. 2006;26(14):5528–5543. doi: 10.1128/MCB.00582-06
[74] Becker A, Zhang P, Allmann L, et al. Poly(ADP-ribosyl)ation of methyl CpG binding protein 2 regulates chromatin structure. J Biol Chem. 2016;291(17):9382. doi: 10.1074/jbc.A115.698357
[75] Rothbauer U, Zolghadr K, Muyldermans S, et al. A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. Mol Cell Proteomics. 2008;7(2):282–289. doi: 10.1074/mcp.M700342-MCP200

[76] Mortusewicz O, Rothbauer U, Cardoso MC, et al. Differential recruitment of DNA ligase I and III to DNA repair sites. Nucleic Acids Res. 2006;34 (12):3523–3532. doi: 10.1093/nar/gkl492
[77] Will CL, Lührmann R. Spliceosome structure and function. Cold Spring Harbor Perspectives In Biology. 2011;3(7): a003707–a003707. doi: 10.1101/cshperspect.a003707
[78] Lee CY, Hubrich D, Varga JK, et al. Systematic discovery of protein interaction interfaces using AlphaFold and experimental validation. Mol Syst Biol. 2024;20(2):75–97. doi: 10.1038/s44320-023-00005-6
[79] Santiago M, Antunes C, Guedes M, et al. TET enzymes and DNA hydroxymethylation in neural development and function — how critical are they? Genomics. 2014;104(5):334–340. doi: 10.1016/j.ygeno.2014.08.018
[80] Razin A, Cedar H. Distribution of 5-methylcytosine in chromatin. Proc Natl Acad Sci USA. 1977;74 (7):2725–2728. doi: 10.1073/pnas.74.7.2725
[81] Kanto S, Grynberg M, Kaneko Y, et al. A variant of Runx2 that differs from the bone isoform in its splicing is expressed in spermatogenic cells. PeerJ. 2016;4: e1862. doi: 10.7717/peerj.1862
[82] Ketchum HC, Suzuki M, Dawlaty MM. Catalytic-dependent and -independent roles of TET3 in the regulation of specific genetic programs during neuroectoderm specification. Commun Biol. 2024;7 (1):415. doi: 10.1038/s42003-024-06120-w
[83] Chrysanthou S, Tang Q, Lee J, et al. The DNA dioxygenase Tet1 regulates H3K27 modification and embryonic stem cell biology independent of its catalytic activity. Nucleic Acids Res. 2022;50(6):3169–3189. doi: 10.1093/nar/gkac089
[84] Dunwell TL, Pfeifer GP. Drosophila genomic methylation: new evidence and new questions. Epigenomics. 2014;6(5):459–461. doi: 10.2217/epi.14.46
[85] Krauss V, Reuter G. DNA methylation in Drosophila–a critical evaluation. Prog Mol Biol Transl Sci. 2011;101:177–191.
[86] Capuano F, Mülleder M, Kok R, et al. Cytosine DNA methylation is found in Drosophila melanogaster but absent in Saccharomyces cerevisiae, schizosaccharomyces pombe, and other yeast species. Anal Chem. 2014;86(8):3697–3702. doi: 10.1021/ac500447w
[87] Wang F, Minakhina S, Tran H, et al. Tet protein function during Drosophila development. PLOS ONE. 2018;13(1): e0190367. doi: 10.1371/journal.pone.0190367
[88] Maniatis T, Reed R. An extensive network of coupling among gene expression machines. Nature. 2002;416 (6880):499–506. doi: 10.1038/416499a
[89] Allemand E, Gattoni R, Bourbon HM. Distinctive features of Drosophila alternative splicing factor RS domain: implication for specific phosphorylation, shuttling, and splicing activation. Mol Cell Biol. 2001;21 (4):1345–1359. doi: 10.1128/MCB.21.4.1345-1359.2001
[90] Prasanth KV, Rajendra TK, Lal AK, et al. Omega speckles - a novel class of nuclear speckles containing hnRnps associated with noncoding hsr-omega RNA in Drosophila. J Cell Sci. 2000;113 Pt 19):3375–3386. doi: 10.1242/jcs.113.19.3485
[91] Hemani Y, Soller M. Mechanisms of Drosophila Dscam mutually exclusive splicing regulation. Biochemical Society Transactions. 2012;40(4):804–809. doi: 10.1042/BST20120060
[92] Ismail JN, Badini S, Frey F, et al. Drosophila tet is expressed in midline glia and is required for proper axonal development. Front Cell Neurosci. 2019;13:252. doi: 10.3389/fncel.2019.00252
[93] Raddatz G, Guzzardo PM, Olova N, et al. Dnmt2-dependent methylomes lack defined DNA methylation patterns. Proc Natl Acad Sci USA. 2013;110 (21):8627–8631. doi: 10.1073/pnas.1306723110
[94] de Almeida Sf, Carmo-Fonseca M, de Almeida SF. The CTD role in cotranscriptional RNA processing and surveillance. FEBS Lett. 2008;582(14):1971–1976. doi: 10.1016/j.febslet.2008.04.019
[95] Tsukamoto T, Hashiguchi N, Janicki SM, et al. Visualization of gene activity in living cells. Nat Cell Biol. 2000;2(12):871–878. doi: 10.1038/35046510
[96] Wang R, Luo Z, He K, et al. Base pairing and structural insights into the 5-formylcytosine in RNA duplex. Nucleic Acids Res. 2016;44(10):4968–4977. doi: 10.1093/nar/gkw235
[97] Behm-Ansmant I, Kashima I, Rehwinkel J, et al. mRNA quality control: an ancient machinery recognizes and degrades mRNAs with nonsense codons. FEBS Lett. 2007;581(15):2845–2853. doi: 10.1016/j.febslet.2007.05.027
[98] Coller J, Parker R. General translational repression by activators of mRNA decapping. Cell. 2005;122 (6):875–886. doi: 10.1016/j.cell.2005.07.012
[99] Li H, Liu H, Zhu D, et al. Biological function molecular pathways and druggability of DNMT2/TRDMT1. Pharmacol Res. 2024;205:107222. doi: 10.1016/j.phrs.2024.107222
[100] Tsumura A, Hayakawa T, Kumaki Y, et al. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells. 2006;11(7):805–814. doi: 10.1111/j.1365-2443.2006.00984.x
[101] Okano M, Bell DW, Haber DA, et al. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999;99 (3):247–257. doi: 10.1016/S0092-8674(00)81656-6
[102] Good CR, Madzo J, Patel B, et al. A novel isoform of TET1 that lacks a CXXC domain is overexpressed in cancer. Nucleic Acids Res. 2017;45(14):8269–8281. doi: 10.1093/nar/gkx435
[103] Rausch C, Zhang P, Casas-Delucchi CS, et al. Cytosine base modifications regulate DNA duplex stability and metabolism. Nucleic Acids Research. 2021;49 (22):12870–12894. doi: 10.1093/nar/gkab509
[104] Stolz P, Mantero AS, Tvardovskiy A, et al. TET1 regulates gene expression and repression of endogenous retroviruses independent of DNA demethylation.

Nucleic Acids Research. 2022;50(15):8491–8511. doi: 10.1093/nar/gkac642

[105] Weyn-Vanhentenryck SM, Feng H, Ustianenko D, et al. Precise temporal regulation of alternative splicing during neural development. Nat Commun. 2018;9 (1):2189. doi: 10.1038/s41467-018-04559-0

[106] Amort T, Rieder D, Wille A, et al. Distinct 5-methylcytosine profiles in poly(A) RNA from mouse embryonic stem cells and brain. Genome Biol. 2017;18(1). doi: 10.1186/s13059-016-1139-1

[107] Hoernes TP, Clementi N, Faserl K, et al. Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. Nucleic Acids Res. 2016;44(2):852–862. doi: 10.1093/nar/gkv1182

[108] Li Q, Li X, Tang H, et al. NSUN2-mediated m5C methylation and METTL3/METTL14-mediated m6A methylation cooperatively enhance p21 translation. J Cell Biochem. 2017;118(9):2587–2598. doi: 10.1002/jcb.25957

[109] Ma H-L, Bizet M, Soares Da Costa C, et al. SRSF2 plays an unexpected role as reader of m5C on mRNA, linking epitranscriptomics to cancer. Molecular Cell. 2023;83 (23):4239–4254.e10. doi: 10.1016/j.molcel.2023.11.003

[110] Yildirim O, Li R, Hung J-H, et al. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell. 2011;147 (7):1498–1510. doi: 10.1016/j.cell.2011.11.054

[111] Sabino JC, de Almeida Mr, Abreu PL, et al. Epigenetic reprogramming by TET enzymes impacts co-transcriptional R-loops. Elife. 2022;11. doi: 10.7554/eLife.69476

Supplementary information

TET dioxygenases localize at splicing speckles and promote RNA splicing

Florian D. Hastert[1,7], Jasmin Weber[1,8], Christina Bauer[2,9], Andreas Zhadan[1], Deepanshu N. D. Singh[3], Thomas C. Dix[3], Roland Arnold[3,4], Sergey Bessonov[6,7], Matthias Soller[3,5], Heinrich Leonhardt[2], M. Cristina Cardoso[1]*, Maria Arroyo[1]*.

*Corresponding author

1. Cell Biology and Epigenetics, Department of Biology, Technical University of Darmstadt, 64287 Darmstadt, Germany.
2. Faculty of Biology and Center for Molecular Biosystems (BioSysM), Human Biology and BioImaging, LMU Munich, Munich 81377, Germany.
3. School of Biosciences, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom.
4. Department of Cancer and Genomics Sciences, School of Medical Sciences, College of Medicine and Health, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom.
5. Division of Molecular and Cellular Function, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, M13 9PT, United Kingdom.
6. Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany.
7. Present address: BioNTech SE, 55131 Mainz, Germany.
8. Present address: LETI Pharma GmbH, Gutenbergstraße 10, 85737 Ismaning, Germany.
9. Present address: BÜHLMANN Laboratories AG, Schönenbuch, Switzerland.

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contacts Maria Arroyo (arroyo.lopez.mc@gmail.com) and M. Cristina Cardoso (cardoso@bio.tu-darmstadt.de).

Supplementary figures and legends

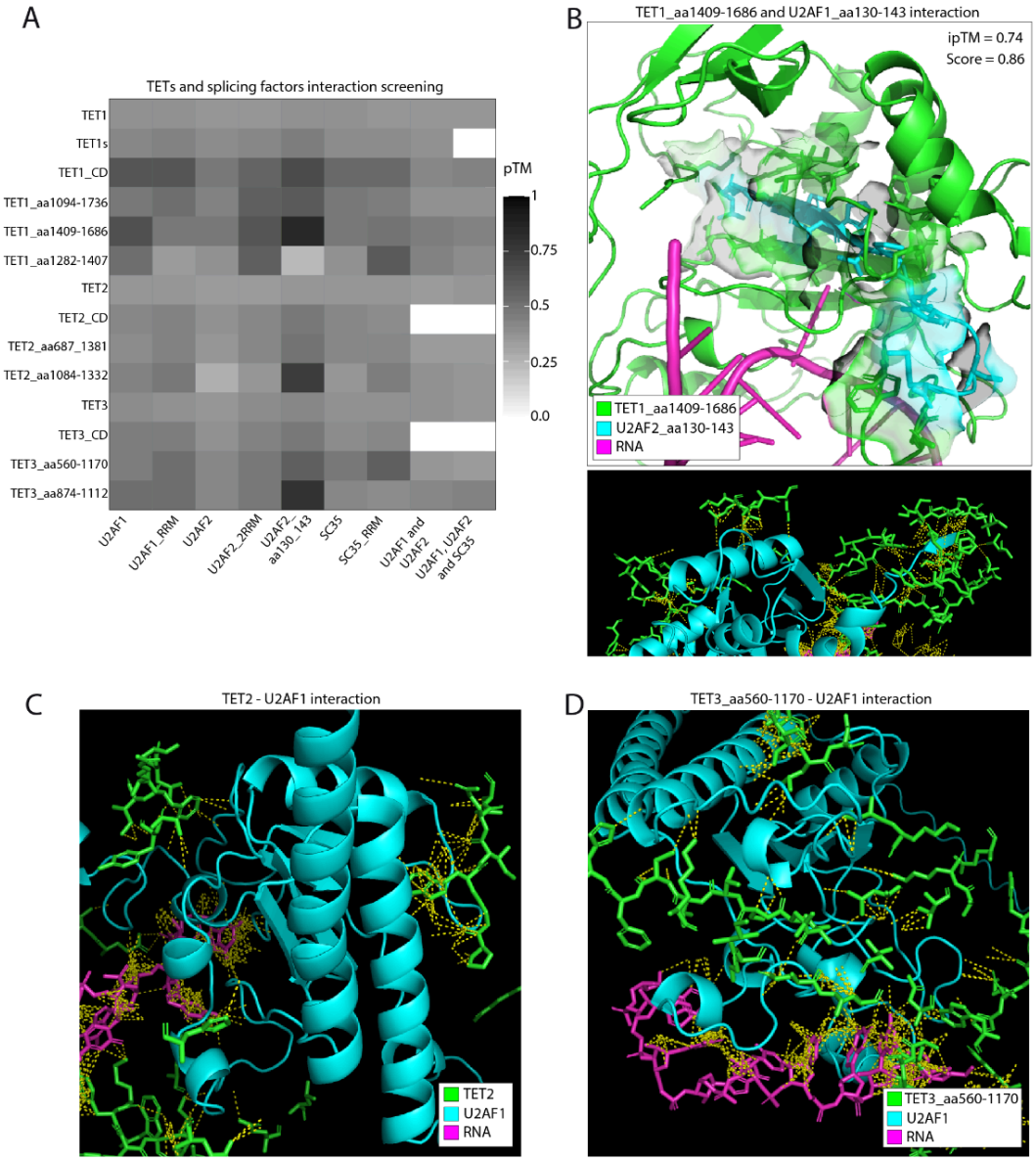


Figure S1. Screening of TET protein interaction with splicing factors using AlphaFold 3. (A) Heatmap showing the pTM score (predicted Template Modeling score) obtained for AF structure models for all the interactions tested between TET proteins and the splicing factors U2AF1, U2AF2, and SC35, together with RNA. Labels on the x and y axes indicate the paired protein fragments for structural modeling. White tiles indicate pairs that were not subjected to structural modeling. Magnification of the highest-scored structural models obtained for **(B)** TET1-U2AF2-RNA, **(C)** TET2-U2AF1-RNA, and **(D)** TET3-U2AF predictions. TET residues within a maximum distance of 5 angstroms to U2AF proteins are shown. All types of contacts between chains at a distance of 4 angstroms are shown as yellow dashed lines.

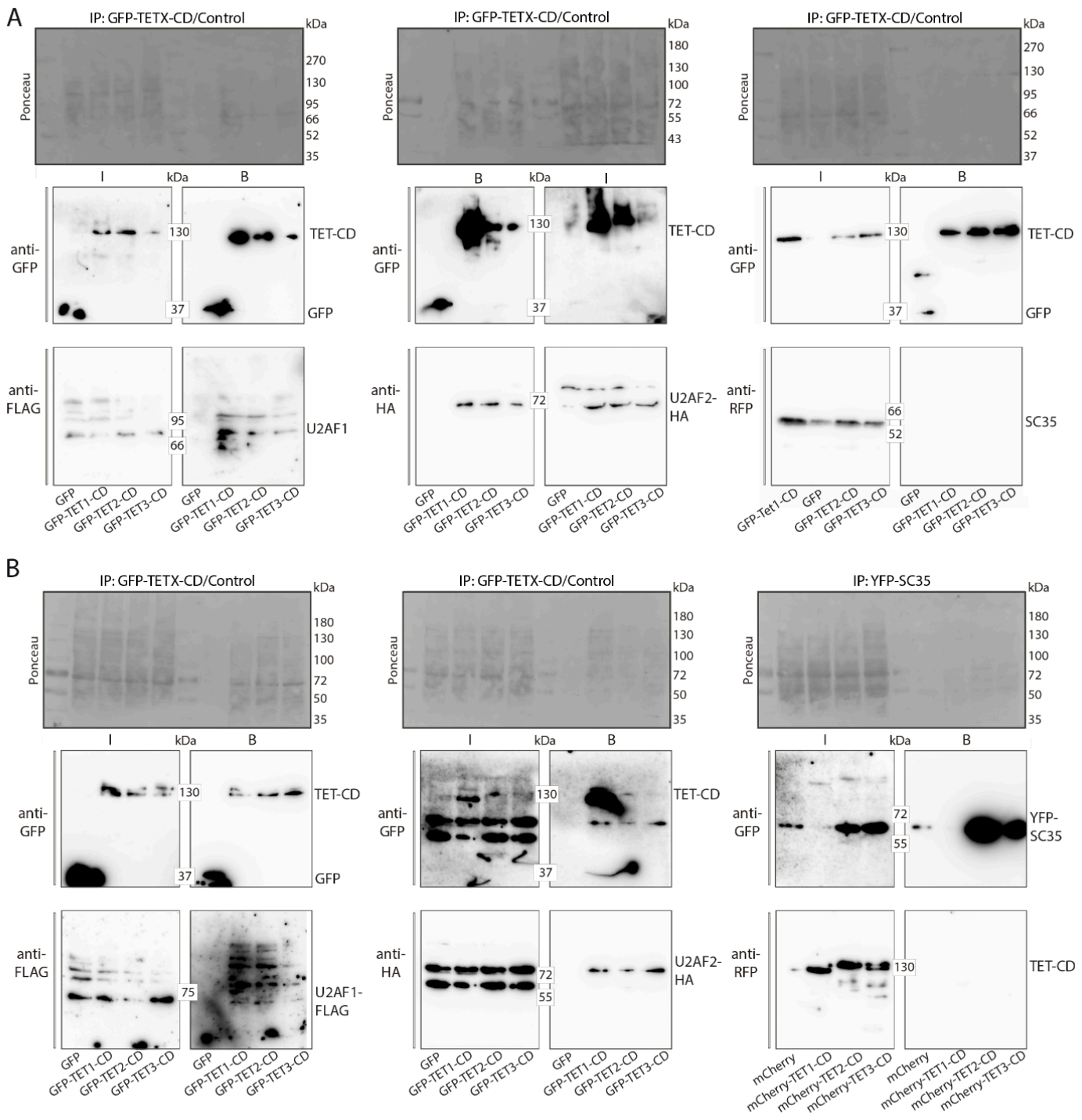


Figure S2. The catalytic domains of TET1, TET2, and TET3 interact with the splicing factors U2AF1 and U2AF2 but not with SC35. (A) Co-immunoprecipitation experiments: HEK cells were transfected with EGFP or EGFP-tagged TET1, TET2, or TET2 catalytic domain, together with the splicing factors U2AF1-FLAG, U2AF2-HA, or mCherry-SC35. Cell extracts were analyzed by immunoprecipitation with immobilized GFP-binding nanobodies, followed by detection with antibodies against GFP or FLAG/HA/mCherry. Input "I" and GFP-binding fraction "B" are shown. **(B)** Co-immunoprecipitation experiments were performed as described in (A) for U2AF1-FLAG and U2AF2-HA. For SC35, HEK cells were transfected with YFP-SC35 and mCherry-tagged TET1, TET2, or TET3 catalytic domains. YFP-SC35 was immunoprecipitated with an immobilized GFP-binding nanobody, followed by detection with antibodies against GFP/RFP.

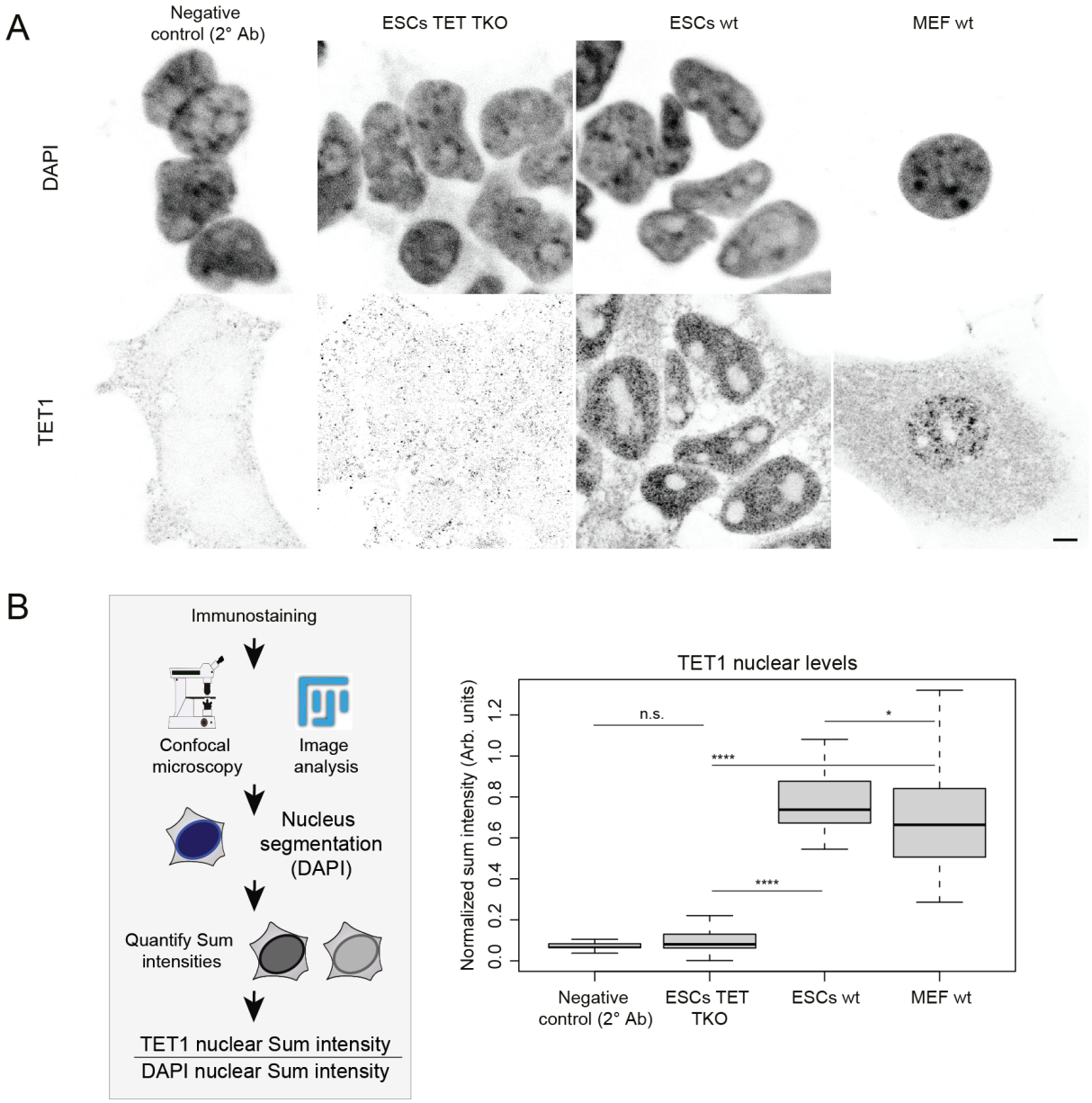


Figure S3. Quantification of TET1 levels in embryonic stem cells (wild type and TET TKO) versus fibroblasts. (A) Representative microscopy images of immunofluorescence against TET1, including a background control (Negative control (2° Ab)), Ab = antibody. This control was performed skipping the step of incubation with primary antibody against TET1. Scale bar: 5 µm. **(B)** On the left, the image analysis pipeline used for the quantification of TET1 immunofluorescence. Cells were imaged using confocal microscopy and quantification was performed using the software Fiji with an automated macro. Nuclei were segmented using the DAPI signal, and the total sum intensity of DAPI and TET1 channels was measured. TET1 sum intensity for each nucleus was normalized to DAPI sum intensity. Normalized values for each condition were plotted using RStudio. On the right, the boxplot shows the results of this quantification. N-numbers (cells): Negative control, 28; TET TKO, 54; ESCs wt, 46; MEF wt, 32. Statistical significance was tested with a paired two-sample Wilcoxon test (n.s., not significant, is given for p-values ≥ 0.05; one star (*) for p-values < 0.05 and ≥ 0.005; two stars (**) is given for values < 0.005 and ≥ 0.0005; three stars (***) is given for values < 0.0005). P-values are shown in Supplementary Table S5.

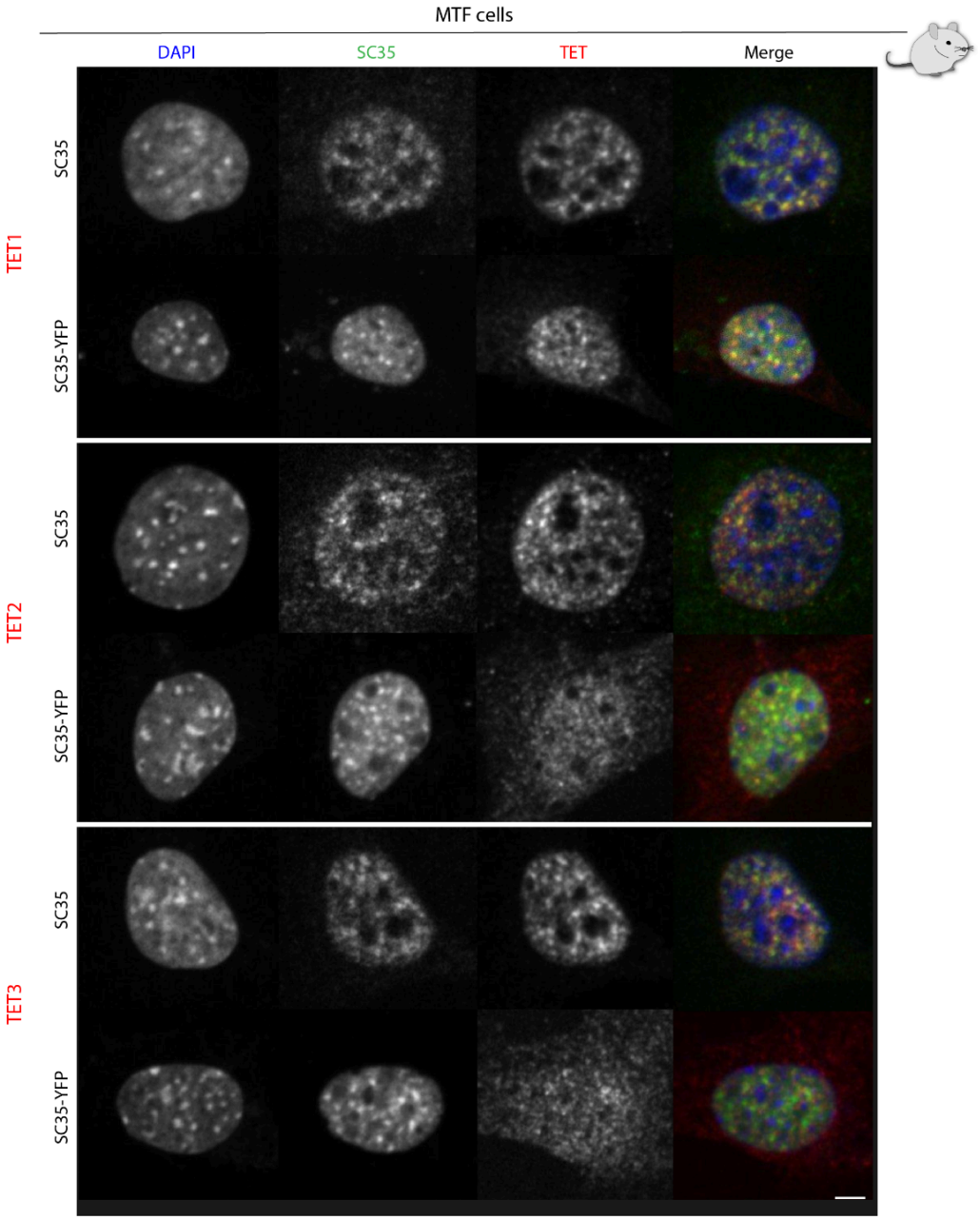


Figure S4. TET1, TET2, and TET3 localization at splicing speckles visualized with SC35 in mouse cells. MTF cells (mouse tail fibroblasts) were immunostained for endogenous TET proteins and SC35 and imaged by confocal microscopy. In addition to endogenous SC35 immunostaining, SC35-YFP was ectopically expressed by transient transfections. Scale bar: 5 μm.

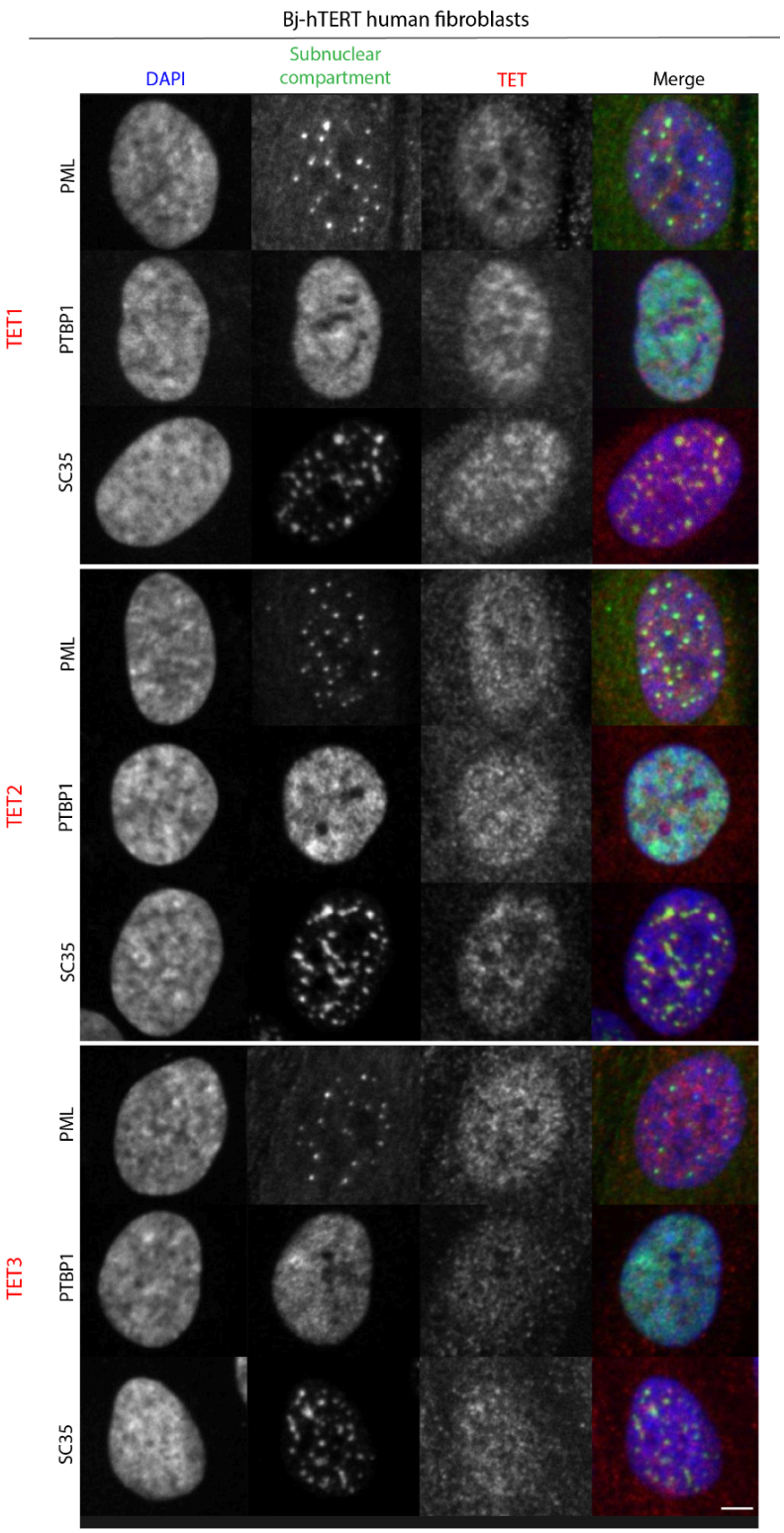


Figure S5. TET1, TET2, and TET3 localization at splicing speckles visualized with SC35 in human cells. Bj-hTERT cells (human foreskin fibroblasts immortalized through the ectopic expression of human telomerase reverse transcriptase (hTERT)) were immunostained for endogenous TET proteins and SC35 and imaged by confocal microscopy. In addition to endogenous SC35 immunostaining, PTBP1 (Polypyrimidine Tract Binding Protein 1) involved in alternative splicing regulation, and PML (Promyelocytic Leukemia Protein) involved in nuclear body formation, were stained. Scale bar: 5 μm.

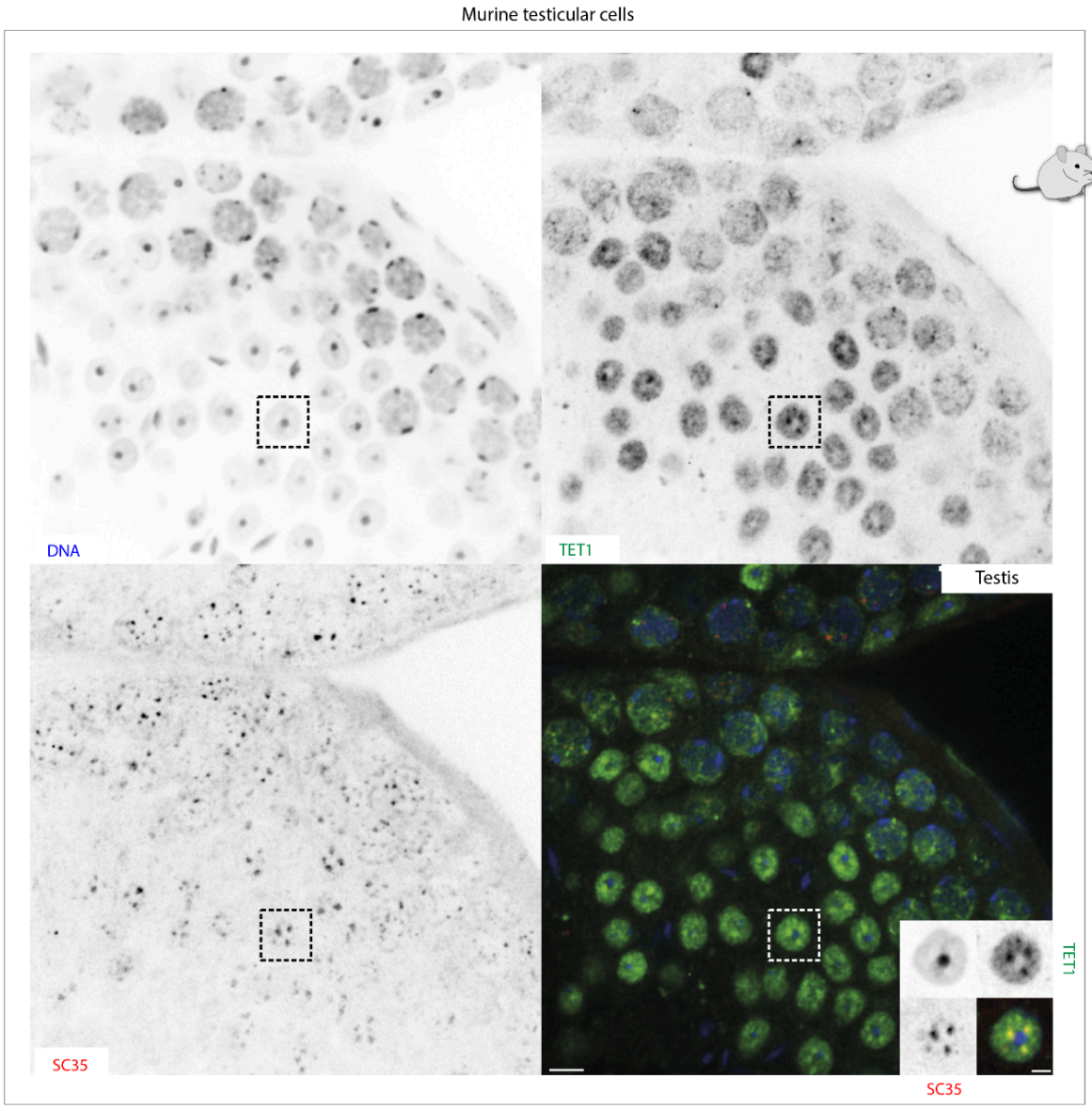


Figure S6. TET1 localizes to SC35-positive speckles in murine testicular cells. Paraffin sections from BALB/c testes were immunostained with antibodies against TET1 and SC35. DNA was counterstained with DAPI. The blowout on the bottom right shows a selected cell and the subnuclear localization of TET1 and SC35 proteins. Stainings for other tissues are shown in Figs. S6 and S7. Scale bar: 5 μm.

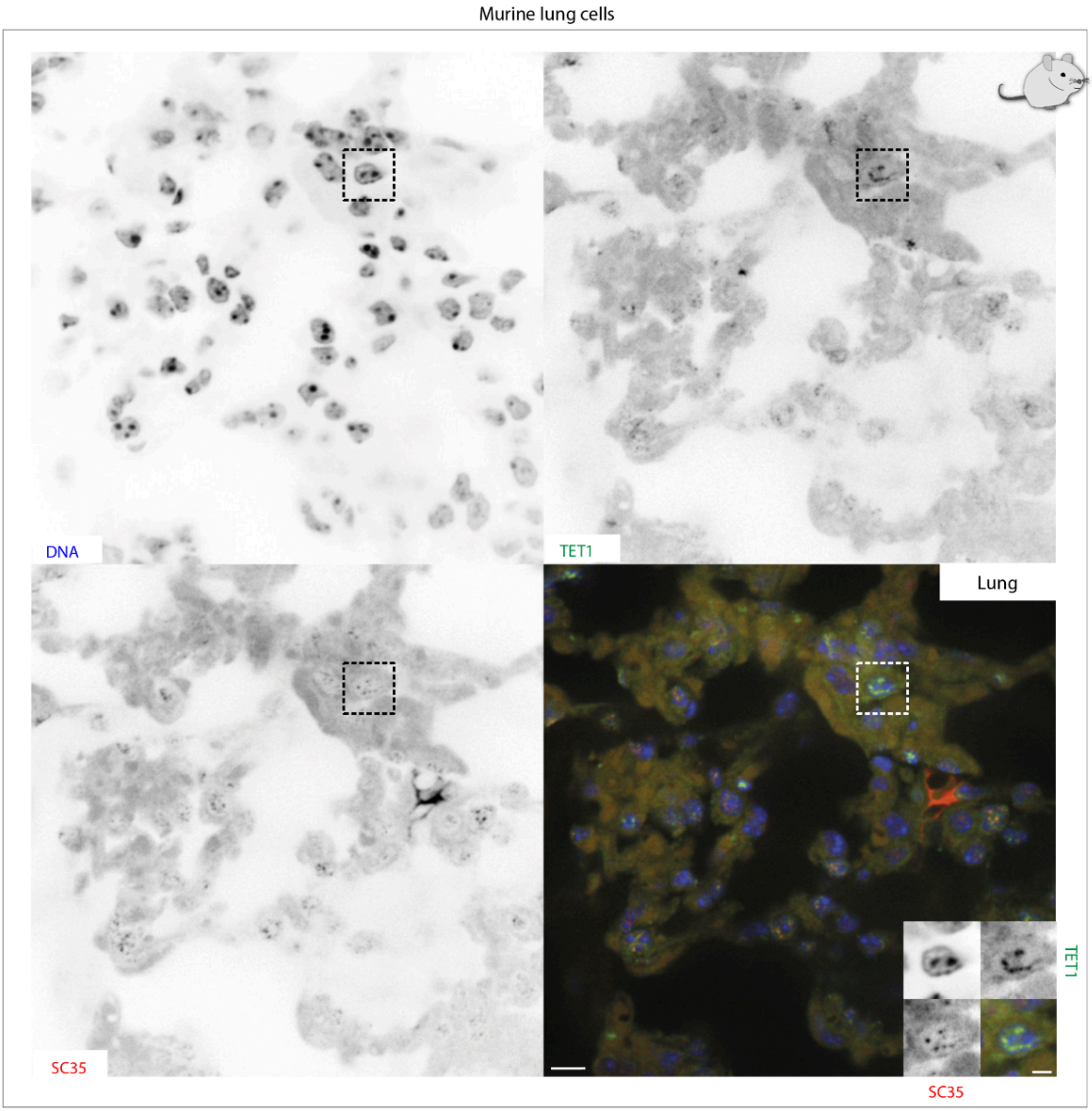


Figure S7. TET1 localizes to SC35-positive speckles in murine lung cells. Paraffin sections were immunostained with antibodies against TET1 and SC35. DNA was counterstained with DAPI. The blowout on the bottom right shows a selected cell and the subnuclear localization of TET1 and SC35 proteins. Scale bar: 5 μm.

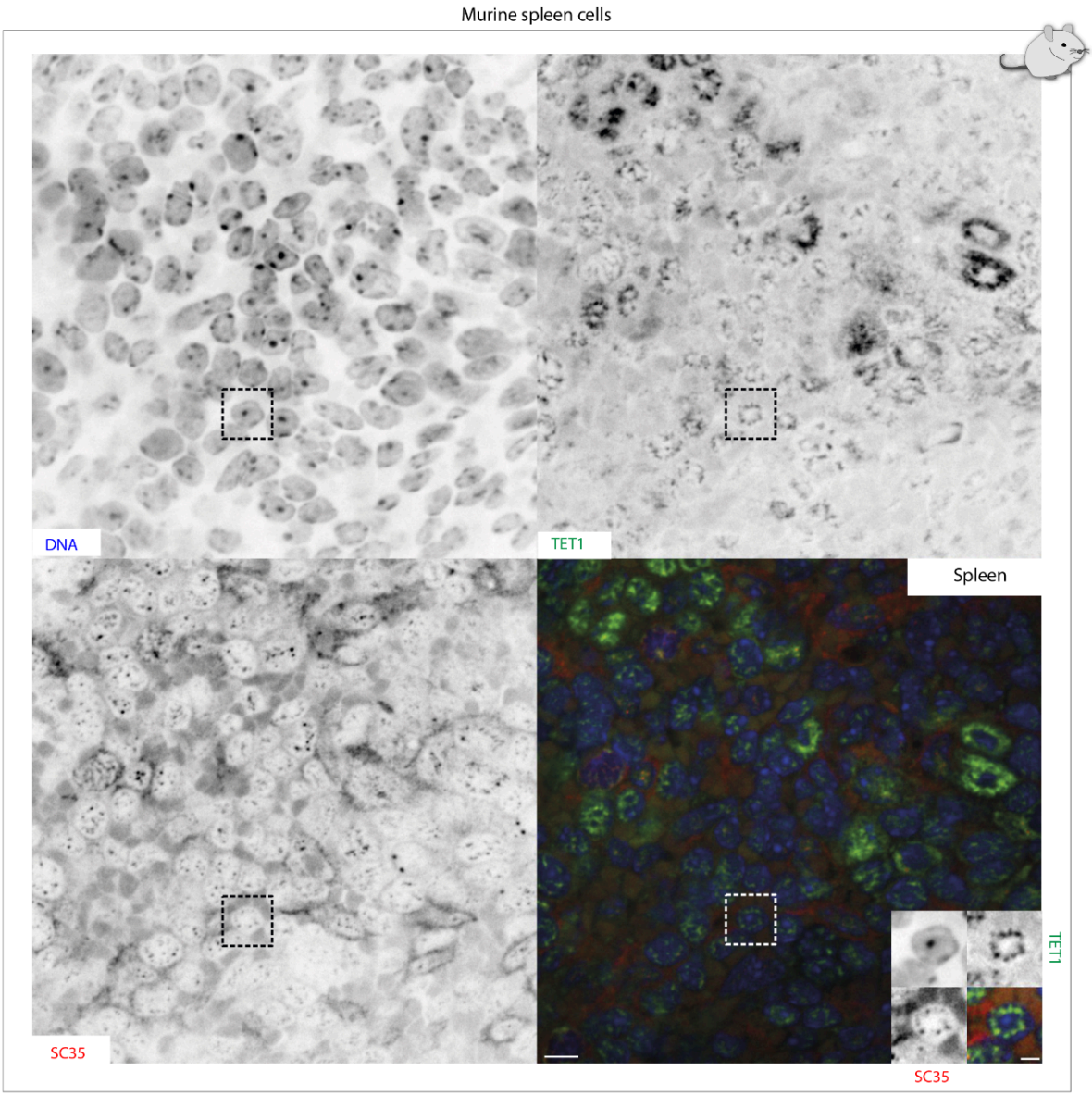


Figure S8. TET1 localizes to SC35-positive speckles in murine spleen cells. Paraffin sections were immunostained with antibodies against TET1 and SC35. DNA was counterstained with DAPI. The blowout on the bottom right shows a selected cell and the subnuclear localization of TET1 and SC35 proteins. Scale bar: 5 µm.

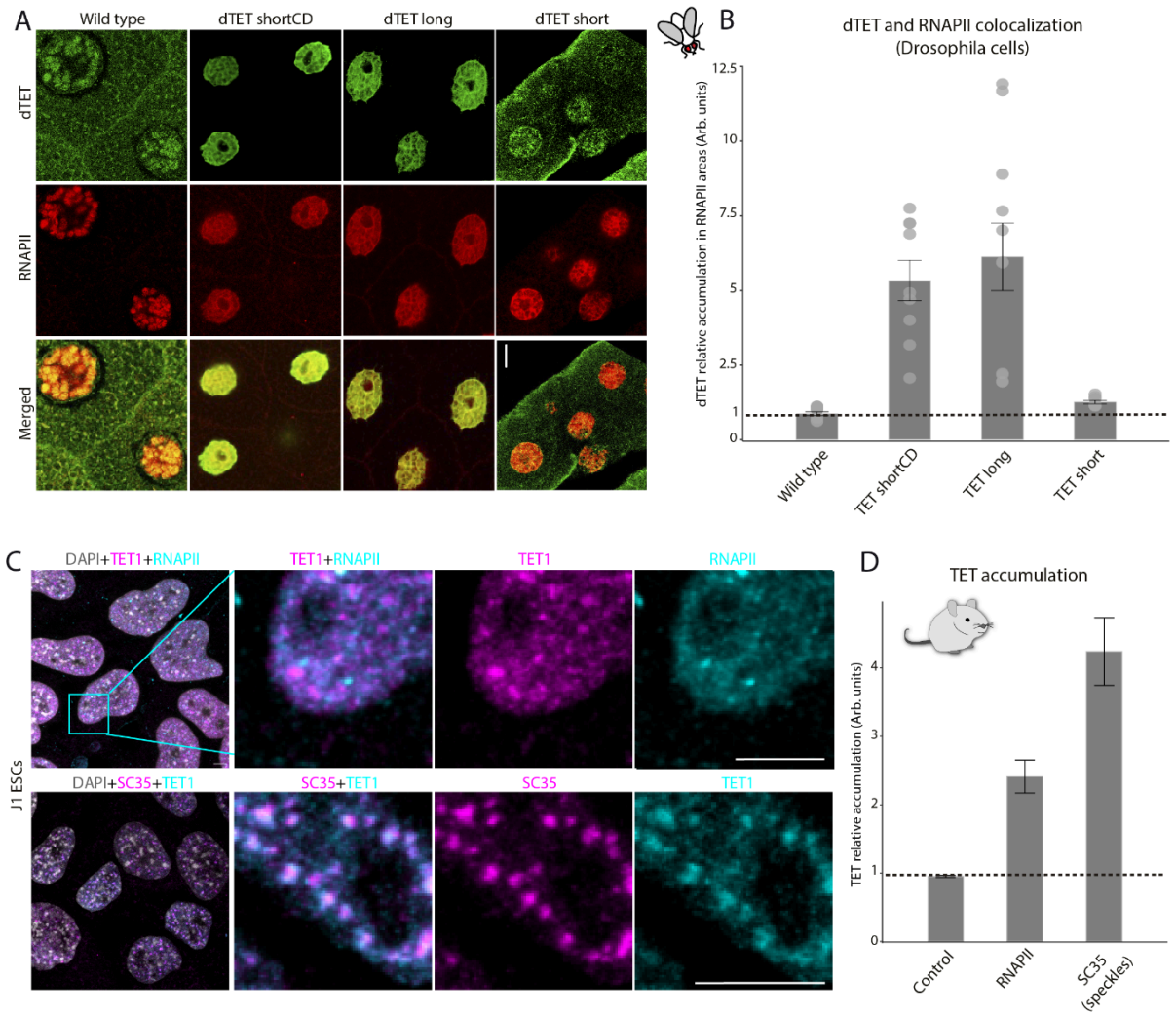


Figure S9. TET1 colocalizes with SC35 and RNAPII in Drosophila and mouse cells. (A) HA-tagged short, long, and short catalytically dead mutant (shortCD) TET isoforms were expressed from UAS inserts by elavC155GAL4 in Drosophila cells. Drosophila third instar larvae salivary glands were stained with anti-HA antibodies to visualize TET-tagged proteins and elongating RNAPII (phosphoSer2) antibodies. Scale bar: 15 µm. **(B)** Colocalization of TET1 and RNAPII was quantified by measuring TET relative accumulation in RNAPII areas as performed in Fig. 3C. Barplot shows the result of this analysis, where the dashed black line separates values above 1, indicating colocalization. **(C)** Representative confocal images of J1 mESCs immunostaining for TET1 and RNAPII or SC35. The magnified areas show both the TET1 and RNAPII/SC35 signal and their overlap. Scale bar: 5 µm. **(D)** Colocalization was quantified by image analysis as described in (B) and is shown as a barplot, with the dashed black line separating values above 1.

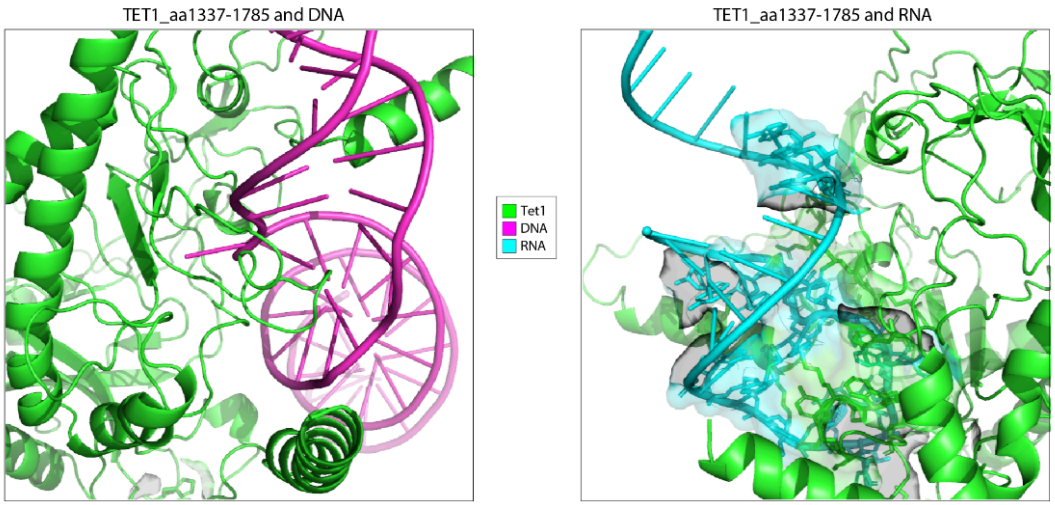


Figure S10. Structural modeling of TET1-CD interaction with DNA and RNA. Magnification of the structural models generated with AlphaFold 3 for TET1-CD and DNA interaction (left) versus RNA (right). As a DNA sequence, a region of the 5' UTR LINE1 promoter (5'-GCGCACCTTCCCTGTAAGAGAGCTTGCCAGCAGAGAGTGCTCTGA-3') was chosen based on previous studies. As an RNA sequence, a short region located in an intron/exon limit was selected (5'-AAAACAUAAAGAAAGGCGUGAGUCUAUGGGA-3'). The sequence used was obtained from the splicing reporters used in this study (spliceable luciferase).

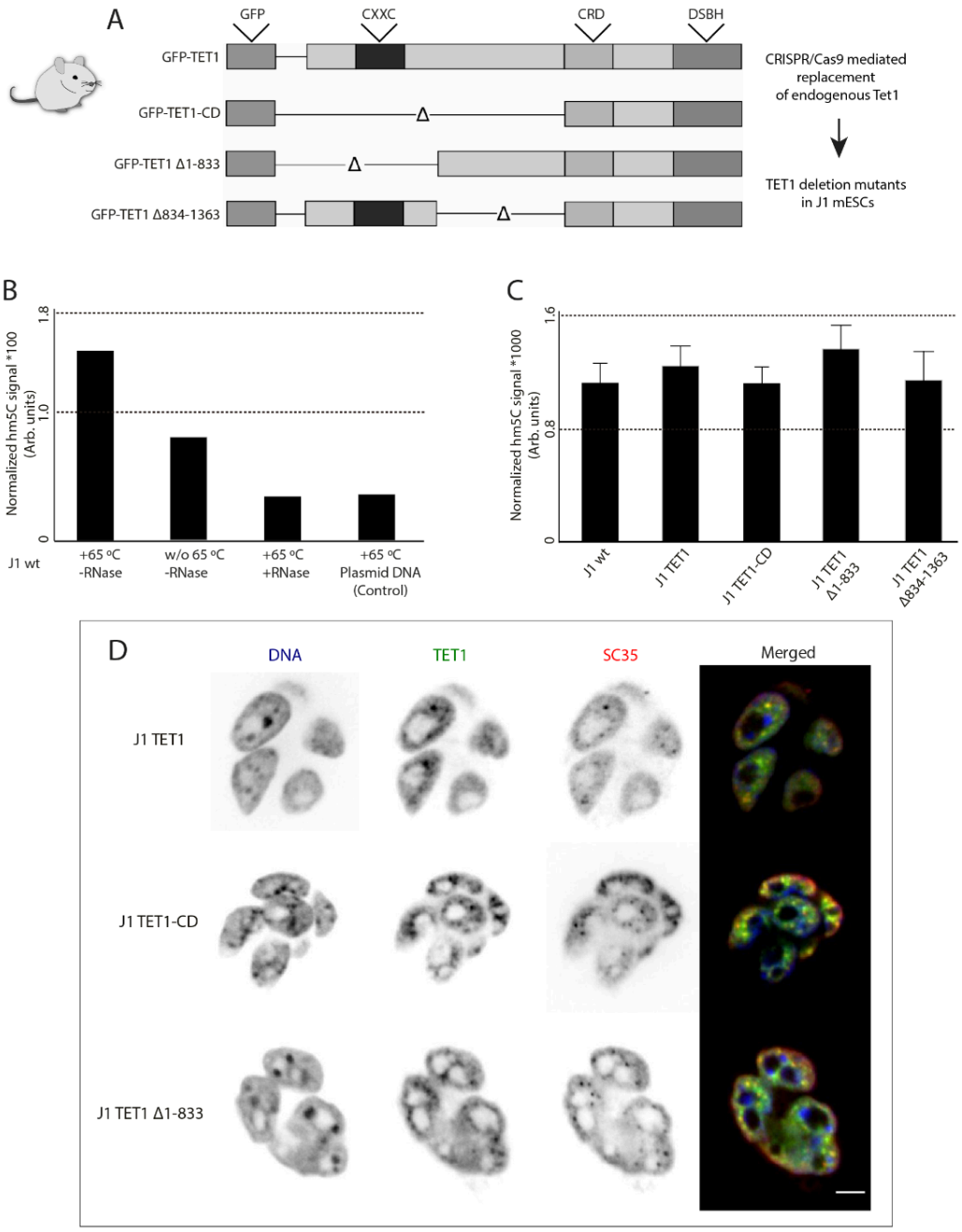


Figure S11. The catalytic domain of TET1 oxidizes m5C to hm5C in RNA secondary structures and localizes in splicing speckles. (A) Scheme showing GFP-tagged TET1 deletion mutant cDNAs used for the genetic edition of J1 wild-type mESCs by CRISPR/Cas9 technology. The full coding sequence of the mutants without introns replaced the TET1 genomic sequence. The cell lines are shown in Supplementary Table S2. **(B)** Total RNA slot blot analysis from J1 wild-type ESCs and image analysis quantification. Barplots showing the quantification for different samples and treatments: RNA boiled at 65°C without RNase A treatment, RNA sample not boiled and not treated with RNase A, RNA sample boiled at 65°C and treated with RNase A, and plasmid DNA digested with EcoRI and HindIII boiled at 65°C as a control. The y-axis shows the hm5C signal in arbitrary units normalized to the methylene blue signal. **(C)** Quantification of slot blot analysis of total RNA isolated from J1 wild-type ESCs and ESCs expressing different TET1 deletion mutants. The y-axis shows the averaged hm5C signal in arbitrary units normalized to the RNA concentration determined by methylene blue intensity. Error bars represent the standard deviation. Experiments were performed as biological triplicates. **(D)** Representative images of immunofluorescence staining for TET1 and SC35 in the different J1 mESCs TET1 deletion mutants. Scale bar: 10 μm.

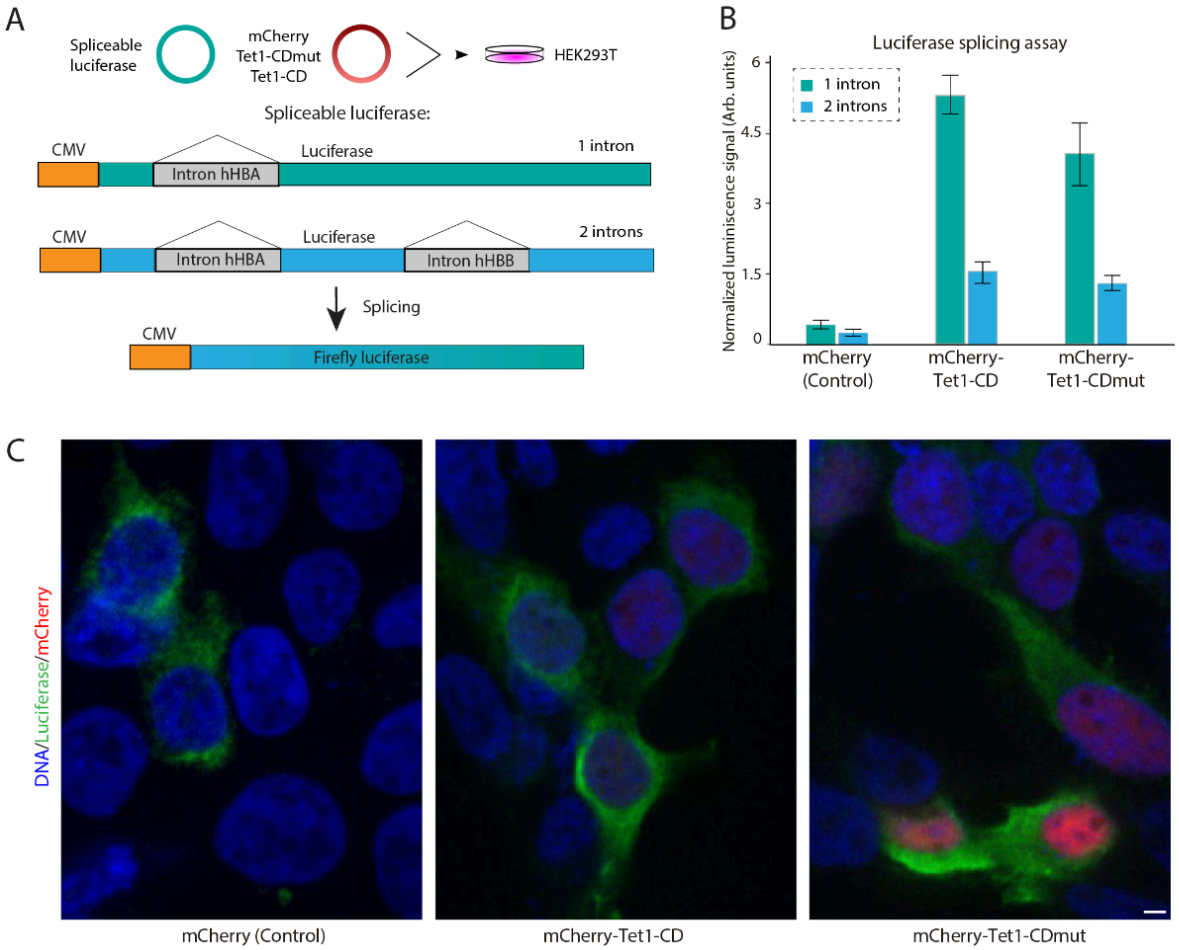


Figure S12. TET1 promotes splicing independent of its catalytic activity. (A) A splicing reporter assay with spliceable firefly luciferase was used to study the role of TET1-CD and TET1-CDm (catalytic dead mutant) in alternative splicing. The diagram shows the pipeline of the experiment and the structure of the luciferase constructs used: spliceable luciferase with two introns (human hemoglobin alpha [hHBA] and human hemoglobin beta [hHBB]) is expressed under the control of a CMV promoter. Alternatively, spliceable luciferase with one intron (hHBA) is also used. **(B)** Barplot showing quantification of the splicing reporter assay by measuring luminescence signal. The results for each construct co-transfected with luciferase genes either containing one (green) or two introns (blue) are shown, and mCherry was used as a control. The quantified luminescence signal in arbitrary units was normalized to the fluorescent signal of mCherry for each condition. Error bars represent the standard deviation. Experiments were performed as biological triplicates. Representative images are shown in **(C)**. Scale bar: 5 µm.

Supplementary tables

Supplementary Table S1: Plasmids

Name	pc number *	Addgene number	Fluorophore/tag	Gene species	Promoter	Reference
pEGFP-N1	0713	-	GFP	*Aequorea victoria*	CMV	Clontech
pGFP-mTet1	2271	-	GFP	*Mus musculus*	CAG	(Frauer et al., 2011)
pGFP-mTet2	2272	-	GFP	*Mus musculus*	CAG	(Bauer et al., 2015)
pGFP-mTet3	2273	-	GFP	*Mus musculus*	CAG	(Liu et al., 2013)
pCherry-C1	2387	-	mCherry	*Discosoma sp.*	CMV	Clontech
pCherry-mTet1-CD	2547	-	mCherry	*Mus musculus*	CAG	(Ludwig et al., 2017)
pCherry-mTet1-CDmut	2815	-	mCherry	*Mus musculus*	CAG	(Ludwig et al., 2017)
pCherry-mTet2-CD	3338	-	mCherry	*Mus musculus*	CAG	(Zhang et al., 2017b, 2017c)
pCherry-mTet3-CD	3339	-	mCherry	*Mus musculus*	CAG	(Ludwig et al., 2017)
pCAG-GFP-Tet1CD	2315	-	GFP	*Mus musculus*	CAG	(Spruijt et al., 2013)
pCAG-GFP-Tet2CD	2316	-	GFP	*Mus musculus*	CAG	(Arroyo et al., 2022; Zhang et al., 2017b, 2017c)
pCAG-GFP-Tet3CD	2309	-	GFP	*Mus musculus*	CAG	(Arroyo et al., 2022; Ludwig et al., 2017)
pU2AF35-FLAG	3967	-	Flag	*Mus musculus*	CMV	Gift from Florian Heyd (Gama-Carvalho et al., 1997; Herdt et al., 2020)
pU2AF65-HA	3968	-	HA	*Mus musculus*	CMV	Gift from Florian Heyd (Gama-Carvalho et al., 1997; Herdt et al., 2020)
pEYFP-SC35	1202	-	YFP	*Mus musculus*	CMV	(Richter et al., 2005)
pmCherry-SC35	3332	235782	mCherry	*Mus musculus*	CMV	This study
pMCP-EGFP-LacI	4555	-	GFP	Artificial sequence	CMV	(Duan et al., 2021)
pms2-PABCP1-mCherry	4573	-	mCherry	*Mus musculus*	CMV	(Duan et al., 2021)
pPABPC1-mCherry	4588	-	mCherry	*Mus musculus*	CMV	(Duan et al., 2021)
pUC18-MINX-M3	3902	-	-	Artificial sequence	T7	Gift from Lührmann Lab (Deckert et al., 2006; Zhang et al., 2017a; Zillmann et al., 1988)
hnRNP-DL-EGFP_ECFP - Splicing reporter NMD	3351	235783	GFP	Synthetic construct	CMV	This study
pWHE200-bg - Splicing reporter luciferase	3335	-	luciferase	Synthetic construct	CMV	Generated by C. Berens lab, provided by Suess Lab (Vogel et al., 2018)
pWHE237-bg - Splicing reporter luciferase	3336	-	luciferase	Synthetic construct	CMV	Generated by C. Berens lab, provided by Suess Lab (Vogel et al., 2018)
pWHE237mod-WT - Splicing reporter luciferase	3337	-	luciferase	Synthetic construct	CMV	Suess Lab (Vogel et al., 2018)

*pc number: plasmid collection number

Supplementary Table S2: Mammalian cell lines

Name	Species	Type	Genotype	Gender	Reference/Source
HEK293-EBNA	*Homo sapiens*	Embryonic kidney	wildtype	female	CVCL_6974 Invitrogen; Paisley, UK
ES v6.5 wt	*Mus musculus*	Embryonic stem cells	wildtype	male	(Dawlaty et al., 2014) Provided by Leonhardt Laboratory.
ES v6.5 Tet TKO	*Mus musculus*	Embryonic stem cells	Tet1,2,3 triple knockout		(Dawlaty et al., 2014; Zhang et al., 2017b) Provided by Leonhardt Laboratory.
ES J1 wt	*Mus musculus*	Embryonic stem cells	wildtype	male	(Li et al., 1992) Provided by Leonhardt Laboratory.

ES J1 Dnmt 1/3a/3b TKO	*Mus musculus*	Embryonic stem cells	Dnmt triple knockout	male	(Okano et al., 1999; Tsumura et al., 2006) Provided by Leonhardt Laboratory.
ES J1 Tet1 Δ 834-1363	*Mus musculus*	Embryonic stem cells	GFP–Tet1Δ834-1363 (endogenous Tet1 deletion)	male	(Mulholland et al., 2015) Provided by Leonhardt Laboratory.
ES J1 Tet1 Δ1- 833	*Mus musculus*	Embryonic stem cells	GFP- Tet1Δ1- 833	male	(Mulholland et al., 2015) Provided by Leonhardt Laboratory.
ES J1 Tet1 CD	*Mus musculus*	Embryonic stem cells	GFP-Tet1CD (endogenous Tet1 deletion)	male	(Mulholland et al., 2015) Provided by Leonhardt Laboratory.
ES J1 Tet1 cDNA F4	*Mus musculus*	Embryonic stem cells	GFP-Tet1 full-length (endogenous Tet1 deletion)	male	(Mulholland et al., 2015) Provided by Leonhardt Laboratory.
MEF W8	*Mus musculus*	Mouse embryonic fibroblasts	wildtype	male	(Peters et al., 2001) (Jenuwein Lab, Freiburg, Germany)
MTF wt line 3	*Mus musculus*	Mouse tail fibroblasts	wildtype	female	(Guy et al., 2001) (Bird Lab, Edinburgh, UK)
BHK clone 2	Mesocricetus auratus	Baby hamster kidney fibroblast	Stable lac-operator array	male	(Tsukamoto et al., 2000) Spector's Laboratory, CSHL, USA
Bj-hTERT	*Homo sapiens*	Skin fibroblast	wildtype	male	(Aladjem and Fanning, 2004; Bodnar et al., 1998) Gift from Mathias Rosenfeldt, University of Wuerzburg, Germany
HeLa	*Homo sapiens*	Cervix epithelium carcinoma	wildtype	female	Stuart Orkin´s lab (HMS), Boston, USA

Supplementary Table S3: Primary and secondary antibodies

Reactivity	Host*	Dilution	Application**	Cat.No. / Clone / ID	Provider / Reference
α-TET1	Rat (mAb)	1:10 / 1:2	IF	Clone 5D8 Clone 4H7	Hybridoma supernatant (Bauer et al., 2015)
α-TET2	Rat (mAb)	1:10 / 1:2	IF	Clone 9F7	Hybridoma supernatant (Bauer et al., 2015)
α-TET3	Rat (mAb)	1:10 / 1:2	IF	Clone 11B6	Hybridoma supernatant (Bauer et al., 2015)
α-GFP	Rat (mAb)	1:1000	WB	Clone 3H9	Chromotek, Planegg-Martinsried, Germany
GFP binder	nanobody	1 mg/mL	coIP/mass spectrometry	-	(Rothbauer et al., 2008)
α-RFP	Rat (mAb)	1:200	WB	Clone 5F8	(Rottach et al., 2008)
α-hemagglutinin (HA)	Mouse (mAb)	1:200	WB/IF	Clone 12CA5	Hybridoma supernatant (Wilson et al., 1984)
α-hemagglutinin (HA)	Rabbit	1:1000	IF Drosophila cells		Sigma-Aldrich Chemie
α-FLAG Tag	Mouse (mAb)	1:200	WB/IF	M2 SLBJ7864V	Sigma-Aldrich Chemie
α-Splicing Factor SC-35	Mouse (mAb)	1:1000	IF	clone S-4045	Sigma-Aldrich Chemie
α-Oct4	rabbit	1:100	IF	ab19857	Abcam, Cambridge, UK
α-RNA polymerase II CTD repeat YSPTSPS (phosphoSer5)	Mouse (mAb)	1:1000	IF	ab 5408	Invitrogen
α-RNA Pol II (phosphoSer2)	Rat (pAb)	1:1000	IF Drosophila cells	clone 3E10, 04-1571	Merck-Millipore
α-5hmC	Rabbit (pAb)	1:250	Slot blot	39769	Active Motif, La Hulpe, Belgium
α-PML (PG-M3)	Mouse (mAb)	1:100	IF	J1904	Santa Cruz Biotechnology
α-PTBP1	Rabbit (mAb)	1:100	IF	ab133734	Abcam, Cambridge, UK

α-Mouse IgG Alexa Fluor®555	Donkey (pAb)	1:250	IF	A31570	Invitrogen, California, USA
α-Mouse IgG Cy3	Donkey (pAb)	1:250	IF	715-166-151	Jackson Immuno Research, Pennsylvania, USA
α-Rat IgG DyLight®488	Donkey (pAb)	1:250	IF	712-546-153	Jackson Immuno Research, Pennsylvania, USA
α-Rat IgG Alexa Fluor ®488	Donkey (pAb)	1:250	IF	712-545-153	Jackson Immuno Research, Pennsylvania, USA
α-Rat IgG Cy5	Donkey (pAb)	1:250	IF	712-175-153	Jackson Immuno Research, Pennsylvania, USA
α-Rat IgG Cy3	Donkey (pAb)	1:250	IF	711-165-152	Jackson Immuno Research, Pennsylvania, USA
α-rabbit IgG (H+L) Alexa Fluor 488-conjugated	Goat (pAb)	1:500	IF	A-11008	Thermo Fisher Scientific, Waltham, MA, USA
α-mouse IgG (H+L) Alexa Fluor 488-conjugated	Goat (pAb)	1:500	IF	A-11001	Thermo Fisher Scientific, Waltham, MA, USA
α-rabbit IgG (H+L) Alexa Fluor 594-conjugated	Goat (pAb)	1:250	IF	R37117	Thermo Fisher Scientific, Waltham, MA, USA
α-Rat IgG Alexa Fluor 647-conjugated	Goat (pAb)	1:250	IF	A-21247	Thermo Fisher Scientific, Waltham, MA, USA
α-Rat IgG (H+L) HRP***-conjugated	Goat (pAb)	1:5000	WB	A9037	Sigma-Aldrich, St Louis, MO, USA
α-Mouse IgG (H+L) HRP***-conjugated	Sheep (pAb)	1:5000	WB	NA931	GE Healthcare, Chicago, Il, USA

*mAb: monoclonal; pAb: polyclonal; **IF: immunofluorescence, WB: western blot; ***HRP: horseradish peroxidase

Supplementary Table S4: Imaging systems

Device	Light Sources	Filters (ex & em [nm])*	Objectives/ Lenses	Detection system	Application
Ultra-View VoX spinning disc on an Inverted Nikon Ti-E microscope; PerkinElmer Life Sciences, UK	Solid state Diode lasers (405 nm, 488 nm, 561 nm, 640 nm)	405/488/568/640 ** 405: 415–475 488: 505–549 561: 580–650 640: 664–754	Oil immersion 60x Plan-Apochromat (NA 1.49) Oil immersion 100x Plan-Apochromat (NA 1.49)	cooled 14-bit Hamamatsu® C9100-50 EMCCD	time-lapse microscopy & confocal Z-stack imaging
Leica SP5 II Confocal point scanner	405 nm Diode Laser 50 mW 488 nm Argon ion laser 458 nm ~5mW 476 nm ~5mW 488 nm ~20mW 496 nm ~5mW 514 nm ~20mW 561 nm DPSS 50 mW 633nm HeNe gas laser 20 mW	DAPI: ex. 420/30 em. 465/20 FITC: ex. 495/15 em. 530/30 Rhod: ex. 570/20 em. 640/40 FITC/Rhod combi filter	HC PL APO 10x / 0.4 CS HCX PL APO 40x / 1.3 oil CS HCX PL APO 63x / 1.4-0.6 oil lambda blue HCX PL APO 100x / 1.44 oil Corr CS	spectral from 400-800 selectivity 0.6-2 nm galvano scanner: up to 1400 Hz resonance scanner: up to 8000 Hz, up to 250 frames per second with 512x512	Time-lapse microscopy & confocal Z-stack imaging
Operetta high content screening microscope; PerkinElmer Life Sciences, UK	Xenon fiber optic Light source, 300 W, 360 – 640 nm continuous spectrum	405: 360-400 & 410–480 488: 460-490 & 500–550 561: 560-580 & 590–640	20x or 40x air (0.45 NA and 0.95 NA) long WD***	14-bit Jenoptik CMOS	high content screening microscopy
Amersham AI600 Imager; GE Healthcare, Chicago, Il, USA	UV transillumination light: 312 nm	EtBr: 312 & 585-625	Large aperture f/0.85 FUJINON™	16-bit Peltier cooled Fujifilm Super CCD	HRP stained blot imaging

*ex: excitation & em: emission, ** dichroic specification, *** WD: working distance

Supplementary Table S5: Statistics

Figure	Sample	n/replicates	Mean	StDev	p-value
1A-C	See Supplementary Data 1				
2C	See Supplementary Data 2				
3A	J1 (TET&DAPI)	10/2	1.221875845	0.18423886	NA
	J1 (TET&SC35)	10/2	2.419255459	0.18133659	
	EB (TET&DAPI)	10/2	0.845532755	0.23053236	
	EB (TET&SC35)	10/2	1.989337709	0.246316896	
	MEF (TET&DAPI)	10/2	0.675793258	0.137883411	
	MEF (TET&SC35)	10/2	1.543048949	0.076502812	
	MTF (TET&DAPI)	10/2	0.720262662	0.154411838	
	MTF (TET&SC35)	10/2	1.786633442	0.147527388	
3B	MTF (TET&DAPI)	10/2	0.432843	0.1934187	NA
	MTF (TET&U2AF1)	10/2	1.769327	0.2681941	
	MTF (TET&DAPI)	10/2	0.419264	0.1876103	
	MTF (TET&U2AF1)	10/2	1.891072	0.2902541	
3D	dTET short	5/2	1.608761001	0.198854875	NA
	dTET long	5/2	1.607218392	0.033577065	
	dTET shortCD	5/2	1.489584822	0.190533916	
	Wild-type (CantonS)	5/2	1.039347709	0.043284972	
4A-D	See Supplementary Data S3-4				
5A	Untreated	13566/3	1	0	NA
	-Triton+RNase	10060/3	0.72934	0.14823	
	+Triton-RNase	11799/3	0.92823	0.00173	
	+Triton+RNase	6203/3	0.26342	0.17329	
5C	(1) Negative control	24/2	1.059673486	0.1633333	(1-2) 8.417e-12
	(2) Positive control	23/2	3.422852525	1.471513136	(1-3) 2.2e-16
	(3) RNA Trap + mRNA mimic	37/2	2.447521656	0.899997807	(2-3) 0.004573
5E	See Supplementary Data 2				
6B	TET Triple KO				NA
	WT+Cherry	13362/2	1.38769884	0.904570142	
	TKO+Cherry	3558/2	1.071902182	0.63286406	
	TKO+TET1CDm	21989/2	5.68629579	2.908236827	
	TKO+TET1CD	23120/2	4.782332972	2.616347675	
	TKO+TET2CD	14513/2	4.657309325	2.958514821	
	TKO+TET3CD	21153/2	4.654218624	2.701217227	
6C	Dnmt Triple KO				NA
	WT+Cherry	8933/2	3.436116618	2.362970511	
	TKO+Cherry	29939/2	2.560897662	1.769923453	
	TKO+TET1CDm	11860/2	4.807776619	2.819110378	
	TKO+TET1CD	15505/2	4.744117214	2.660777855	
	TKO+TET2CD	2852/2	4.173808474	2.816790199	
	TKO+TET3CD	14303/2	4.870746262	2.652516561	
S3	Negative control (2° Ab)	28/1	0.069563414	0.020278148	(1-2) 0.05591
	ESCs TET TKO	54/1	0.155127646	0.24296153	(2-4) 5.138e-12
	ESCs wt	46/1	0.806650298	0.216138888	(2-3) 1.492e-15
	MEF wt	32/1	0.67726321	0.258059176	(3-4) 0.01911
S9B	Wild type	10/2	0.869797722	0.208917705	NA
	TET shortCD	9/2	5.335722866	2.04129062	
	TET long	11/2	6.123324018	3.757580461	
	TET short	8/2	1.257700419	0.162348096	
S9D	Control	12/2	0.954414887	0.047514536	NA
	RNAPII	14/2	2.41385661	0.873790948	
	SC35 (speckles)	23/2	4.242922308	2.208899833	
S11B	+65 ºC -RNase	/1	1.5396	-	NA
	w/o 65 ºC -RNase	/1	0.8293	-	
	+65 ºC +RNase	/1	0.2945	-	
	+65 ºC Plasmid DNA (control)	/1	0.3021	-	
S11C	J1 wt	/3	1.1256	0.2156	NA
	J1 TET1	/3	1.3573	0.2573	
	J1 TET1-CD	/3	1.1102	0.1985	
	J1 TET1-Δ1-833	/3	1.3814	0.3012	
	J1 TET1-Δ8341363	/3	1.0991	0.3928	
S12B	mCherry (1 intron)	>190377/3	0.2301	0.1134	NA
	mCherry (2 introns)	>89627/3	0.1683	0.1021	
	mCherry-TET1-CD (1 intron)	>168140/3	5.2638	0.7632	
	mCherry-TET1-CD (2 introns)	>70563/3	1.6112	0.3198	
	mCherry-TET1-CDmut (1 intron)	>81505/3	3.6183	0.9856	
	mCherry-TET1-CDmut (2 introns)	>161862/3	1.2196	0.2659	

Supplementary References

Aladjem MI, Fanning E. 2004. The replicon revisited: an old model learns new tricks in metazoan chromosomes. *EMBO Rep* **5**:686–691. doi:10.1038/sj.embor.7400185

Arroyo M, Hastert FD, Zhadan A, Schelter F, Zimbelmann S, Rausch C, Ludwig AK, Carell T, Cardoso MC. 2022. Isoform-specific and ubiquitination dependent recruitment of Tet1 to replicating heterochromatin modulates methylcytosine oxidation. *Nat Commun* **13**:5173. doi:10.1038/s41467-022-32799-8

Bauer C, Göbel K, Nagaraj N, Colantuoni C, Wang M, Müller U, Kremmer E, Rottach A, Leonhardt H. 2015. Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosamine transferase (OGT). *J Biol Chem* **290**:4801–4812. doi:10.1074/jbc.M114.605881

Bodnar AG, Ouellette M, Frolkis M, Holt SE, Chiu CP, Morin GB, Harley CB, Shay JW, Lichtsteiner S, Wright WE. 1998. Extension of life-span by introduction of telomerase into normal human cells. *Science* **279**:349–352. doi:10.1126/science.279.5349.349

Dawlaty MM, Breiling A, Le T, Barrasa MI, Raddatz G, Gao Q, Powell BE, Cheng AW, Faull KF, Lyko F, Jaenisch R. 2014. Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev Cell* **29**:102–111. doi:10.1016/j.devcel.2014.03.003

Deckert J, Hartmuth K, Boehringer D, Behzadnia N, Will CL, Kastner B, Stark H, Urlaub H, Lührmann R. 2006. Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol Cell Biol* **26**:5528–5543. doi:10.1128/MCB.00582-06

Duan N, Arroyo M, Deng W, Cardoso MC, Leonhardt H. 2021. Visualization and characterization of RNA-protein interactions in living cells. *Nucleic Acids Res* **49**:e107. doi:10.1093/nar/gkab614

Frauer C, Rottach A, Meilinger D, Bultmann S, Fellinger K, Hasenöder S, Wang M, Qin W, Söding J, Spada F, Leonhardt H. 2011. Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PLoS ONE* **6**:e16627. doi:10.1371/journal.pone.0016627

Gama-Carvalho M, Krauss RD, Chiang L, Valcárcel J, Green MR, Carmo-Fonseca M. 1997. Targeting of U2AF65 to sites of active splicing in the nucleus. *J Cell Biol* **137**:975–987. doi:10.1083/jcb.137.5.975

Guy J, Hendrich B, Holmes M, Martin JE, Bird A. 2001. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat Genet* **27**:322–326. doi:10.1038/85899

Herdt O, Reich S, Medenbach J, Timmermann B, Olofsson D, Preußner M, Heyd F. 2020. The zinc finger domains in U2AF26 and U2AF35 have diverse functionalities including a role in controlling translation. *RNA Biol* **17**:843–856. doi:10.1080/15476286.2020.1732701

Liu X, Gao Q, Li P, Zhao Q, Zhang J, Li J, Koseki H, Wong J. 2013. UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nat Commun* **4**:1563. doi:10.1038/ncomms2562

Li E, Bestor TH, Jaenisch R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**:915–926. doi:10.1016/0092-8674(92)90611-f

Ludwig AK, Zhang P, Hastert FD, Meyer S, Rausch C, Herce HD, Müller U, Lehmkuhl A, Hellmann I, Trummer C, Storm C, Leonhardt H, Cardoso MC. 2017. Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise. *Nucleic Acids Res* **45**:2438–2457. doi:10.1093/nar/gkw1197

Mulholland CB, Smets M, Schmidtmann E, Leidescher S, Markaki Y, Hofweber M, Qin W, Manzo M, Kremmer E, Thanisch K, Bauer C, Rombaut P, Herzog F, Leonhardt H, Bultmann S. 2015. A modular open platform for systematic functional studies under physiological conditions. *Nucleic Acids Res* **43**:e112. doi:10.1093/nar/gkv550

Okano M, Bell DW, Haber DA, Li E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**:247–257. doi:10.1016/s0092-8674(00)81656-6

Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schöfer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A, Opravil S, Doyle M, Sibilia M, Jenuwein T. 2001. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**:323–337. doi:10.1016/s0092-8674(01)00542-6

Richter K, Reichenzeller M, Görisch SM, Schmidt U, Scheuermann MO, Herrmann H, Lichter P. 2005. Characterization of a nuclear compartment shared by nuclear bodies applying ectopic protein expression and correlative light and electron microscopy. *Exp Cell Res* **303**:128–137. doi:10.1016/j.yexcr.2004.09.022

Rothbauer U, Zolghadr K, Muyldermans S, Schepers A, Cardoso MC, Leonhardt H. 2008. A versatile

nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Mol Cell Proteomics* **7**:282–289. doi:10.1074/mcp.M700342-MCP200

Rottach A, Kremmer E, Nowak D, Leonhardt H, Cardoso MC. 2008. Generation and characterization of a rat monoclonal antibody specific for multiple red fluorescent proteins. *Hybridoma (Larchmt)* **27**:337–343. doi:10.1089/hyb.2008.0031

Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PWTC, Bauer C, Münzel M, Wagner M, Müller M, Khan F, Eberl HC, Mensinga A, Brinkman AB, Lephikov K, Müller U, Walter J, Boelens R, van Ingen H, Leonhardt H, Carell T, Vermeulen M. 2013. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**:1146–1159. doi:10.1016/j.cell.2013.02.004

Tsukamoto T, Hashiguchi N, Janicki SM, Tumbar T, Belmont AS, Spector DL. 2000. Visualization of gene activity in living cells. *Nat Cell Biol* **2**:871–878. doi:10.1038/35046510

Tsumura A, Hayakawa T, Kumaki Y, Takebayashi S, Sakaue M, Matsuoka C, Shimotohno K, Ishikawa F, Li E, Ueda HR, Nakayama J, Okano M. 2006. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**:805–814. doi:10.1111/j.1365-2443.2006.00984.x

Vogel M, Weigand JE, Kluge B, Grez M, Suess B. 2018. A small, portable RNA device for the control of exon skipping in mammalian cells. *Nucleic Acids Res* **46**:e48. doi:10.1093/nar/gky062

Wilson IA, Niman HL, Houghten RA, Cherenson AR, Connolly ML, Lerner RA. 1984. The structure of an antigenic determinant in a protein. *Cell* **37**:767–778. doi:10.1016/0092-8674(84)90412-4

Zhang P, Hastert FD, Ludwig AK, Breitwieser K, Hofstätter M, Cardoso MC. 2017a. DNA base flipping analytical pipeline. *Biol Methods Protoc* **2**:bpx010. doi:10.1093/biomethods/bpx010

Zhang P, Ludwig AK, Hastert FD, Rausch C, Lehmkuhl A, Hellmann I, Smets M, Leonhardt H, Cardoso MC. 2017b. L1 retrotransposition is activated by Ten-eleven-translocation protein 1 and repressed by methyl-CpG binding proteins. *Nucleus* **8**:548–562. doi:10.1080/19491034.2017.1330238

Zhang P, Rausch C, Hastert FD, Boneva B, Filatova A, Patil SJ, Nuber UA, Gao Y, Zhao X, Cardoso MC. 2017c. Methyl-CpG binding domain protein 1 regulates localization and activity of Tet1 in a CXXC3 domain-dependent manner. *Nucleic Acids Res* **45**:7118–7136. doi:10.1093/nar/gkx281

Zillmann M, Zapp ML, Berget SM. 1988. Gel electrophoretic isolation of splicing complexes containing U1 small nuclear ribonucleoprotein particles. *Mol Cell Biol* **8**:814–821. doi:10.1128/mcb.8.2.814-821.1988