TET2 chemically modifies tRNAs and regulates tRNA fragment levels

Chongsheng He[1,2,3,†], **Julianna Bozler**[2,3], **Kevin A. Janssen**[4,5], **Jeremy E. Wilusz**[5], **Benjamin A. Garcia**[2,4,5], **Andrea J. Schorn**[6], **Roberto Bonasio**[2,3,†]

[1]College of Biology, Hunan Key Laboratory of Plant Functional Genomics and Developmental Regulation, Hunan University, Changsha, 410082 Hunan, P.R. China

[2]Epigenetics Institute, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

[3]Department of Cell and Developmental Biology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

[4]Graduate Group in Biochemistry and Molecular Biophysics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

[5]Department of Biochemistry and Biophysics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

[6]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

Abstract

The Ten-eleven Translocation 2 (TET2) protein, which oxidizes 5-methylcytosine in DNA, can also bind RNA; however, the targets and function of TET2–RNA interactions *in vivo* are not fully understood. Using stringent affinity tags introduced at the *Tet2* locus, we purified and sequenced TET2-crosslinked RNAs from mouse embryonic stem cells (mESCs) and found a high enrichment for tRNAs. RNA immunoprecipitation with an antibody against 5-hydroxymethylcytosine (hm5C) recovered tRNAs that overlapped with those bound to TET2 in cells. Mass spectrometry analyses revealed that TET2 is necessary and sufficient for the deposition of the hm5C modification on tRNA. *Tet2* knockout in mESCs affected the levels of several small noncoding RNAs originating from TET2-bound tRNAs that were enriched by hm5C immunoprecipitation. Thus, our results suggest a novel function of TET2 in promoting the conversion of 5-methylcytosine to hm5C on tRNA and regulating the processing or stability of different classes of tRNA fragments.

† To whom correspondence should be addressed: chongshenghe@outlook.com; roberto@bonasiolab.org.

INTRODUCTION

Chemical modifications of RNA bases are increasingly being appreciated as essential regulators of various steps of RNA metabolism, including splicing, export, and degradation[1,2]. Several recent studies have shown the biological function of N^6-methyladenosine (m6A) modification on messenger RNAs (mRNAs), non-coding RNAs (ncRNAs), and microRNAs[3–5] (miRNAs). The presence of m6A can affect RNA stability, secondary structure, RNA translation and nuclear export. Besides m6A, 5-methylcytosine (m5C) and 5-hydroxymethylcytosine (hm5C) modifications have also been detected on mRNAs, transfer RNAs (tRNAs), and ribosomal RNAs[6–10] (rRNAs). Those two modifications were reported to control translational efficiency of mRNAs[7] and stability of tRNAs[11–14]. However, the biological roles of hm5C in RNA have not been fully defined.

Ten-eleven Translocation (TET) proteins were discovered as dioxygenases that function on DNA to convert 5-methylcytosine, through subsequent oxidation reactions, into 5-hydroxymethyl-, 5-formyl-, and 5-carboxyl-cytosine, which is eventually excised by thymine-DNA glycosylase (TDG) resulting in net DNA demethylation[15]. Because of their involvement in the first steps of DNA demethylation, TET proteins are believed to be key epigenetic regulators involved in a variety of biological processes including embryonic development, neurogenesis, and immunity[16–18].

Although for several years the catalytic activity of TET proteins was thought to be restricted to DNA, recent studies have begun to unravel their potential role in modifying RNA bases. All three mammalian TET proteins, TET1, TET2, and TET3, can catalyze the oxidation of m5C on RNA *in vitro*[19,20]. In *Drosophila*, dTET is responsible for hm5C modification of mRNAs, which was reported to affect translation efficiency[7]. Furthermore, TET2 can target m5C in mRNAs and endogenous retrovirus (ERV) transcripts in mammalian cells[18,21], and it has been proposed to regulate protein–RNA interactions and RNA stability either via the loss of m5C or the deposition of hm5C on the target transcript[18,21]. Generally, it is now believed that TET proteins can use both DNA and RNA as substrate, however, the full spectrum of RNAs, in particular noncoding transcripts targeted by TET proteins *in vivo*, remain poorly characterized.

Originating from tRNAs, tRNA fragments (tRFs) are abundant small RNAs found in many organisms[22]. Previously overlooked as a potential sequencing artefact, their biological roles have come into sharper focus in recent years, as several studies showed their role in regulating translation[14,23–25], transposable elements[26] (TEs), and epigenetic inheritance[27–30]. Two main classes of tRFs from mature tRNAs are currently recognized: 5' tRFs and 3' tRFs, based on whether they contain 5' or 3' sequences[22]. 5' tRFs can be further divided into tRF5a, tRF5b, tRF5c and 5tiR according to the position of the cleavage site and the chemical nature of their ends, whereas 3' tRFs can be classified as tRF3a, tRF3b and 3tiR.

We previously reported that TET2 is an RNA-binding protein *in vivo*[31]. Using photocrosslinking and mass spectrometry in mouse embryonic stem cells (mESCs), we mapped the RNA binding region of TET2 to a peptide adjacent to the catalytic domain,

suggesting that RNA might be an enzymatic substrate of TET2 *in vivo*[31]. To further dissect the mechanistic and functional relationship between TET2 and RNAs, we generated epitope-tagged alleles of *Tet2* in mESCs using CRISPR/Cas9 technology. Using these cell lines, we confirmed that TET2 binds to RNA *in vivo* and found tRNA as its major substrate, with these interactions resulting in the deposition of hm5C. We also found that loss of TET2 results in altered levels of tRFs.

Our findings suggest an intriguing molecular link between tRNA processing and RNA modifications mediated by TET2, which might have broad implications in development and disease, given the regulatory roles of tRFs in metabolism, cancer, and epigenetics.

RESULTS

TET2 binds to RNA in mESCs

TET2 is a methylcytosine dioxygenase better known for its role in DNA demethylation[15], however, we previously recovered TET2 in a screen for novel putative RNA-binding proteins, validated its RNA-binding activity in HEK293 cells, and mapped it to a peptide adjacent to the catalytic site[31]. As our original screen was performed in mESCs and given that TET2 is highly expressed in these cells[32] as well as blastula-stage embryos[33], we decided to study the biological role of TET2–RNA interactions in mESCs.

We sought to validate the RNA-binding activity of TET2 in mESCs using protein–RNA crosslinking after 4-thiouridine (4SU) treatment followed by immunoprecipitation[34] (henceforth CLIP for simplicity). Our initial attempts using commercial antibodies were limited by excessive background signal; therefore, we decided to insert two affinity purification tags (6xHis and HA, Fig. 1A) at the N terminus of the endogenous *Tet2* locus using CRISPR/Cas9-mediated genome editing (see methods for details).

We obtained several mESC clones with the correct insertion for TET2 as determined by PCR screening and selected two clones that were confirmed by Sanger sequencing (Extended Data Fig. 1A). In TET2-tagged clones (#65 and #75), a protein of the correct size for His-HA-TET2 was specifically immunoprecipitated with an HA antibody (Fig. 1B).

We strategically chose these two epitope tags as they would allow us to employ very stringent washing conditions during the CLIP protocol, which is notoriously prone to background noise, especially when applied to non-canonical RNA binding proteins[35,36]. Specifically, 6xHis-tagged proteins can be purified on nickel- or cobalt-coupled beads in denaturing conditions, up to 8 M urea[37]. We developed our own protocol for these CLIP experiments, incorporating a tandem 6xHis and HA affinity purification and borrowing principles from the previously published PAR-CLIP[34], eCLIP[36], and irCLIP[38] strategies. Overall, our customized CLIP protocol minimizes background and avoids radioactive labeling (Fig. 1C, see methods for details).

We detected clear signal from RNA crosslinked to TET2 *in vivo* in mESCs (Fig. 1D). The fact that the fluorescent signal increased when mESCs were pulsed with 4SU proved that it originated from crosslinked protein–RNA species. The two independently isolated clones

gave comparable results (Fig. 1D) and quantification of the fluorescent signal further confirmed that it was 4SU-dependent, demonstrating the presence of crosslinked RNA (Extended Data Fig. 1B). Therefore, by using our optimized CLIP strategy, we found that endogenous TET2 binds to RNA in mESCs.

Genome-wide identification of TET2-bound RNAs *in vivo*

Next, we sought to identify the RNAs bound to TET2 in mESCs. We redesigned the eCLIP library construction and sequencing strategy[36] to simplify the protocol and the downstream bioinformatic analyses (methods) and combined it with our tandem affinity CLIP strategy. We constructed libraries from two independent pull-downs for each of the two separate clones, resulting in four total biological replicates. As originally described in the eCLIP method[36], we also constructed libraries for size-matched inputs to control for background signal (Extended Data Fig. 2A).

We sequenced input and CLIP libraries close to saturation and obtained 10–70 million raw reads comprising 1–2 million unique molecular identifiers (UMIs) per CLIP library and 4–19 million UMIs per input library (Supplementary Table 1). Our modified CLIP protocol yielded highly reproducible results, as shown by strong correlations between all input samples, and, separately, between all TET2 CLIP samples, including different biological replicates and different 6xHis–HA-tagged mESC clones (Fig. 2A). Analyzing the transcripts identified in the TET2 CLIP samples compared to the input, we noticed that tRNAs were strongly enriched in the TET2-associated fraction (Fig. 2B, red dots), suggesting an affinity of TET2 for this class of ncRNAs. In fact, tRNAs constituted the most enriched class of RNAs bound to TET2 when considering both the fraction enriched (Extended Data Fig. 2B) and the odds ratio for the enrichment (Fig. 2C).

Because CLIP-seq recovers small RNA fragments surrounding the crosslinking site, conventional RNA-seq count analyses, which assume coverage of the entire transcript by the sequencing reads, might be inappropriate and biased toward the identification of small RNAs (such as tRNAs) in the enriched fraction. To avoid potential biases and better define transcripts bound to TET2 based on local enrichment of sequencing reads, we adapted an algorithm previously used for MeRIP-seq[39] (see methods for details) and found 4,027 regions with an FDR-corrected *P*-value for CLIP enrichment lower than 10^{-5}, which we defined as TET2 CLIP peaks. These peaks mapped to most classes of annotated transcripts and were particularly frequent among tRNAs, with 65% of annotated tRNAs containing at least one CLIP peak (Extended Data Fig. 2C). In fact, tRNAs were by far the most significantly enriched class of annotated RNAs containing TET2 CLIP-seq peak (Fig. 2D), confirming our previous conclusions based on transcript-level analyses.

The only known TET-family enzyme in *Drosophila*, dTET, was reported to bind to mRNAs[7]. Mammalian TET2 was shown to bind directly to mRNAs in human bone marrow-derived macrophages[18] and indirectly to ERVs in mESCs[21]. Our CLIP-seq in mESCs extends these findings adding various categories of ncRNAs to the TET2-bound transcriptome, including, most notably, tRNAs.

5-hydroxymethycytosine is enriched at TET2-binding sites on tRNAs

In addition to targeting DNA, TET2 can also function as a dioxygenase that uses RNA as a substrate[18,19,21]; however, whether this activity targets small RNAs (sRNAs), including tRNAs, *in vivo* has not been investigated. To answer this question, we sought to map the distribution of the hm5C modification on sRNAs by performing RNA immunoprecipitation (RIP) and sequencing using an antibody against hm5C[21] followed by an sRNA library construction protocol, obtaining 3–10 million mappable reads per sample (Supplementary Table 2). We performed the RIP in two biological replicates and sequenced it along with input controls. Libraries obtained from control IgG immunoprecipitation could not be properly analyzed due to limited library complexity. Similar to our results for CLIP-seq the biological replicates for the input and hm5C RIP samples were highly consistent and clustered with each other (Extended Data Fig. 3).

To identify RNAs significantly enriched, we utilized the same peak-calling algorithm based on a window-based Fisher test as above[39]. Overall, we detected 5,943 RNA regions with an FDR-corrected P-value for anti-hm5C RIP enrichment lower than 10^{-5}, contained in 1,846 unique transcripts. Among the four classes of sRNAs (< 200 nts in length) that were detectable in the deep sequencing data, tRNAs were the only class significantly ($P = 4.1 \times 10^{-25}$, Fisher's exact test) enriched, with 21% (76/359) of all detectable tRNAs containing one or more hm5C peak (Fig. 3A). Moreover, 59 out of these 76 hm5C-containing tRNAs were also recovered by TET2 CLIP (Fig. 3A–B, Supplementary Table 3), a proportion significantly higher than expected by chance ($P < 0.01$, hypergeometric distribution). The overlap remained significant even when relaxing the stringent FDR cutoff.

Next, we analyzed the spatial relationship between the CLIP and anti-hm5C RIP signal. Metaplot analyses showed that the peak of the anti-hm5C RIP signal on tRNAs coincided with the peak of the CLIP signal (Fig. 3C, left), whereas no discernible anti-hm5C RIP signal was observed at TET2-bound snRNAs (Fig. 3C, middle). As an additional control we inspected the signal on rRNAs, which were also enriched in the TET2 CLIP experiments (see Fig. 2D). Overall the anti-hm5C RIP signal peaked on rRNAs at the same position as the CLIP signal, but the relative enrichment (counts per million) was much lower than for the tRNA class (Fig. 3C, right). Visual inspection of heatmaps for all four sRNA classes confirmed the global observations on class enrichment and the spatial colocalization of CLIP and hm5C on tRNAs but not snRNAs, snoRNAs, or miRNAs (Fig. 3D).

In conclusion, the enrichment of tRNAs after hm5C RIP and the spatial colocalization of the TET2 CLIP and hm5C RIP signals are consistent with the notion that TET2 binds to tRNAs *in vivo* and catalyzes the oxidation of m5C to hm5C.

TET2 is the major enzyme responsible for hm5C deposition on tRNA in mESCs

To confirm the presence of hm5C on tRNAs with an antibody-independent method, we utilized RNA mass spectrometry. Using column- ("crude small RNA fraction") and gel-based size fractionation ("tRNAs", Extended Data Fig. 4A) and synthetic standards (Extended Data Fig. 4B), we quantified the abundance of hm5C on RNA obtained from ESCs. Although hm5C was detectable in the total RNA fraction, it was greatly enriched both

in the sRNA and tRNA fraction, 11- and 12-fold, respectively (Fig. 4A). Given that tRNAs account for 5–10% of total RNA and for most of the RNA in the sRNA fraction, this result suggests that the majority of hm5C on RNA originates from tRNAs.

Next, we sought to determine whether TET enzymes were required for the deposition of hm5C on tRNAs *in vivo*. We generated a loss-of-function allele for *Tet2* by inserting a STOP cassette a the 5' of the locus using the same targeting strategy described above (Extended Data Fig. 4C; see methods for details). Purified tRNA from *Tet2* KO mESCs contained 41% less hm5C than the corresponding fraction from WT cells (Fig. 4B, $P = 2 \times 10^{-4}$), showing that TET2 is necessary for proper hm5C modification of tRNAs *in vivo*. Similar results were obtained when analyzing the crude small RNA fraction (Extended Data Fig. 4D). Removal of the two remaining TET enzymes, TET1 and TET3, in a *Tet1/2/3* triple KO (tKO) mESC line[40] resulted in a smaller but significant additional loss (13%, $P = 0.03$) of hm5C (Fig. 4B).

To confirm the catalytic function of TET2 on tRNAs *in vivo*, we transiently overexpressed a 100 kDa C-terminal fragment of TET2 encompassing the catalytic domain (WT) or the same fragment containing key mutations that inactivate the enzyme (HxD[41]) in *Tet1/2/3* tKO mESCs. Despite relatively low levels of transgenic expression (Fig. 4C), TET2 WT but not the catalytically inactive mutant caused a significant increase in total hm5C detected by mass spec on tRNAs (Fig. 4D). Finally, incubation of recombinant TET2 catalytic domain (Fig. 4E) with total tRNAs purified from *Tet1/2/3* tKO mESCs *in vitro* resulted in an increased abundance of hm5C as detected by mass spectrometry (Fig. 4F). Together these two experiments demonstrate that TET2 is sufficient for hm5C deposition on tRNAs *in vivo* and *in vitro.*

The observation that tRNAs from *Tet1/2/3* tKO cells are depleted for hm5C is consistent with a previous report[42], and our mass spectrometry analyses now reveal that TET2 is the major enzyme responsible for hm5C deposition on tRNAs in mESCs.

TET proteins regulate processing of tRNAs into tRFs

Having shown that TET2 deposits hm5C on tRNAs *in vivo*, we wanted to determine the function of this activity. Intriguingly the post-transcriptional RNA modification status of tRNA, in particular the presence of m5C, has been shown to regulate the levels of various tRNA-derived sRNAs[11–14]. In fact, deletion of *Nsun2*, the major source of m5C on tRNAs[43] results in increased number of sRNA reads mapping to the 5' of a subset of affected tRNAs and a concomitant decrease of signal from their 3' portion[14]. These tRNA-derived sRNAs are abundant in most organisms and are broadly classified into tRF5 and tRF3 according to their origin in the full-length tRNA molecule[22] (5' and 3', respectively). Depending on their size, 3' tRFs can be further divided into tRF3a (17–19 nts) or tRF3b (22 nts), whereas 5' tRFs comprise species 18–35 nts long[22]. Among other functions[22,44], tRFs inhibit reverse transcription and translation of ERVs[26] and have been proposed to serve as the vehicle for transgenerational epigenetic inheritance[27,28].

We hypothesized that TET-mediated conversion of m5C to hm5C on tRNAs or tRFs might affect their processing or stability. To test this hypothesis, we compared tRNA-mapping

reads from small RNA-seq libraries in control WT mESCs and TET2-deficient mESCs in three biological replicates. Global analyses revealed a decreased coverage for the 5' portion of mature tRNAs and a concomitant increase in signal over their 3' end (Fig. 5A). This conclusion was confirmed when only analyzing reads in the proper size range for each tRF category (Fig. 5B); that is, 28–35 nts for tRF5s, 17–19 nts for tRF3a, or 22 nts for tRF3b (for tRF3's we only considered reads containing the three non-template CCA nucleotides at their very end). These changes are in the opposite direction as those observed in *Nsun2*$^{-/-}$ cells[14], consistent with the enzymatic opposing roles of TET2 (decreases m5C) and NSUN2 (increases m5C).

These global changes were reflected in asymmetric effects of TET loss on the different classes of tRFs originating from individual tRNAs. Among the tRF3a class, 32 distinct tRF sequences were found at significantly higher levels (adjusted *P*-value < 0.1) in *Tet2*$^{-/-}$ cells, whereas only 6 tRF3a's were decreased (Fig. 5C, middle). Similarly, 32 distinct tRF3b's were present at significantly higher levels in *Tet2*$^{-/-}$ cells and none showed decreased abundance (Fig. 5C, right). Importantly, these results were confirmed in independent experiments, also comprising three biological replicates per genotype, where we used a different ESC line (E14) as WT control (Extended Data Fig. 5).

We asked whether TET2 binding *in vivo* (CLIP) or hm5C levels (RIP) could predict the effects on the respective tRFs as measured in the *Tet2*$^{-/-}$ mESCs. Indeed a substantial portion of tRF3a's upregulated in *Tet2*$^{-/-}$ cells overlapped with TET2 CLIP peaks (Fig. 5D) as well as anti-hm5C RIP peaks (Fig. 5E) suggesting a link between TET2 binding and changes in tRF levels.

To investigate a connection between m5C and TET2-mediated regulation of tRFs, we reanalyzed a high-quality RNA BS-seq obtained in mESCs by Legrand *et al*[10]. We confirmed known tRNA methylation patterns, including a small number of tRNAs that contained m5C at position 38, which is placed by DNMT2, and a large number of tRNAs containing m5C at NSUN2 methylation sites (Extended Data Fig. 6A). Consistent with a link between TET2 and NSUN2 in regulating tRFs, we found that tRF3a's and tRF3b's upregulated in *Tet2*$^{-/-}$ cells displayed substantial overlap with tRNAs highly methylated (> 75% m5C) at NSUN2-dependent positions (Extended Data Fig. 6B). Analyses of individual tRNAs confirmed these genome-wide observations: several known NSUN2 target tRNAs, including the well-characterized tRNA-LeuCAA[13], were the source of dysregulated tRFs in *Tet2*$^{-/-}$ cells, bound to TET2, and were enriched by hm5C RIP (Extended Data Fig. 6C–K).

Our results suggest that the opposing activities of NSUN2 and TET2 on tRNAs regulate the processing or stability of tRFs.

DISCUSSION

TET2 binds RNA in mESCs

We previously discovered an unexpected RNA-binding activity within the catalytic region of TET2, which we validated in HEK293[31]. Here, we expanded on the functional relevance of our previous finding by showing that TET2 binds to different types of RNA in mESCs. We

developed a stringent CLIP protocol leveraging tandem affinity purification on dual epitope-tagged proteins modified at their endogenous locus with genome editing. In the course of our experiments, we also tested TET1 for RNA binding but could not detect a clear signal by CLIP (data not shown). We did not test the RNA binding capability of TET3, as it is not expressed in mESCs, but it will be interesting to explore this in future studies, as TET3 appears to have dedicated functions in the brain[45,46].

Conversion of m5C to hm5C on RNA by TET2

Although we and others reported the RNA-binding activity of TET2, a comprehensive characterization of RNAs that interact with this enzyme in mESCs has been lacking. We found that TET2 interacted with a diverse array of noncoding transcripts including those originating from tRNA, snRNA, and rRNA genes. Although TET2 can bind different types of RNA, we chose to focus on tRNAs, because they were by far the most enriched category according to our CLIP results (Fig. 2). Considering that TET2 can oxidize m5C to hm5C on RNA[19,20], it is reasonable to hypothesize that transcripts with high levels of m5C might be primary targets for TET2 binding. Indeed it has been known for decades that tRNAs contain m5C[47,48], which might explain why they are the most abundant class of RNA bound by TET2. We provide three lines of evidence to support our conclusion that tRNAs are physiological substrates for TET2: 1) anti-hm5C RIP enriches preferentially tRNAs bound by TET2 (Fig. 3); 2) hm5C levels on tRNA are reduced in *Tet2*[−/−] mESCs as determined by direct quantification via mass spectrometry and partially rescued *in vivo* by transfection of the TET2 catalytic domain (Fig. 4); 3) TET2 catalyzes the conversion of m5C to hm5C on purified tRNA *in vitro* (Fig. 4). We also observed strong TET2 CLIP signal from snRNAs, but they did not contain measurable hm5C according to our RIP-seq results, suggesting the possibility that TET2 might have non-enzymatic functions in association with snRNAs. Another intriguing possibility is that snRNAs might carry other methylated bases (e.g. m6A) that might be also subjected to oxidation by TET2. We found limited residual hm5C in tRNAs from *Tet1/2/3* tKO cells, which may be caused by non-catalytic oxidation, although we cannot exclude that other enzymes might contribute to this process *in vivo*.

All mass spectrometry measurements were performed on total tRNA and therefore cannot determine if specific tRNAs contain more hm5C than others, which is a limitation of this study. We can however conclude that full-length tRNAs do contain this modification, as our gel-purification strategy excluded potential tRFs based on size. Although we speculate that the hm5C decrease in TET-deficient cells results in stabilization and retention of m5C, the predicted changes in m5C would be smaller than our measurement error because of the low ratio of hm5C to m5C. Development of direct RNA-sequencing by mass spectrometry or the adaptation of specific m5C and hm5C sequencing techniques developed for DNA might help answer these questions in the future.

Function of hm5C on RNA

In mRNA, hm5C modification can affect translational efficiency and secondary structure[18] while in ERV transcripts it affects RNA stability[21]. However, a function of hm5C in ncRNAs has not been reported. In this study, we found that deletion of *Tet2* and lower levels of hm5C on tRNAs correlate with increased global levels of tRF3a and tRF3b and decreased tRF5

(Fig. 5). This is consistent with opposite effects reported after deletion of the methyltransferase *Nsun2*[14]. Together, these observations lead us to propose that hm5C regulate the biogenesis of some tRFs either directly, or indirectly by replacing m5C[12–14]. Although we did not investigate the downstream consequences of TET2-induced changes in tRF levels or their targets, the functions of these small RNAs have been demonstrated in various contexts[22,26–28,44]. Nonetheless, the effects of TET-mediated modification of tRNAs on tRF levels are small and future research is required to establish the extent of their biological relevance.

We note that without direct experimental manipulation of hm5C levels on tRNAs, alternative interpretations for our findings are possible. For example, because the same catalytic domain of TET2 is involved in both RNA and DNA modification, we cannot exclude that some of the effects observed on tRFs are indirectly due to the activity of TET2 on DNA. We also note that, in other contexts, m5C has been reported to regulate translation and it remains possible that TET2-mediated modifications on tRNAs might also affect translation independently or via the regulation of tRFs, which are interesting avenues to pursue in the future.

Previous studies found that tRFs could be processed by DICER[49,50], however, a majority of 3' and 5' tRFs are still present in *Dicer*[−/−] mESCs, indicating the involvement of alternative endonucleases[51]. Other studies have shown that tRNA halves or tiRs are generated by the RNase A-like ribonuclease angiogenin under stress condition[52] and that this cleavage could be inhibited by m5C modification[11,14], but this process was not observed under resting conditions. In both cases, only specific tRNAs were subject to processing. Thus, it remains unknown how the full spectrum of tRFs is generated *in vivo* and what molecular pathways regulate this process. Our results provide one clue: generation of some tRFs might be regulated by hm5C modification. The RNA methyltransferases DNMT2 and NSUN2 add m5C to tRNAs during maturation thereby regulating their stability[11,14,25,53–59]. 5'-hydroxymethylation of tRNAs by TET2 could be acting in an antagonistic way. We propose that the balance between m5C and hm5C is important to regulate tRNA stability or perhaps the efficiency of loading into currently unknown processing complexes.

Conclusions and outlook

We have characterized with unbiased genome-wide methods the TET2-bound transcriptome in mESCs as well as the distribution of hm5C on sRNAs. Only tRNAs had high levels of bound TET2 as well as hm5C modification, consistent with previous studies reporting the presence of m5C on this RNA species[10,47,48]. We found that the presence of TET2 represses tRF3a and tRF3b generation and enhances tRF5 levels, thereby altering tRNA fate. It is still unknown how the majority of tRFs are processed from tRNAs and how tRF levels are regulated. We propose that TET2 regulates tRF levels by modifying the m5C/hm5C ratio on tRNAs, which has important implications for the role of this key epigenetic modifier in development and disease.

METHODS

Cell culture

Mouse ESCs C57BL/6 X 129/Sv (WT) and *Tet1/2/3* tKO[40] were obtained from Marisa Bartolomei (University of Pennsylvania). E14Tg2A.4 (E14) mESCs were obtained from Danny Reinberg (New York University). *Tet2* KO were generated for this study (see below). All mESCs were cultured on gelatin-coated dishes in KnockOut DMEM (Thermo Fisher) supplemented with 15% FBS (Thermo Fisher), 100 mM MEM nonessential amino acids (Sigma), 0.1 mM 2-mercaptoethanol (Sigma), 1 mM L-glutamine (Invitrogen), 0.5% Penicillin Streptomycin (Sigma), 100 U/mL leukemia inhibitory factor (LIF) (Chemicon), 3 μM CHIR99021 (Millipore) and 1 μM PD0325901 (Millipore)[60]. Cells were routinely tested for mycoplasma and when needed genotype was confirmed by qPCR or WB.

Construction of TET2 knock-in and KO mESCs

For CLIP experiments, CRISPR/Cas9 knock-in E14 lines were generated by transient transfection of the relevant single guide RNA (sgRNA) constructs and single-stranded DNA donor (Supplementary Table 4) followed by selection with 1 μg/mL puromycin (Invitrogen). Clones were screened by PCR and the genotype of the positive clones was confirmed by Sanger sequencing.

We employed a similar strategy construct *Tet2* KO mESCs. The transfection was carried out using the same sgRNAs and a different donor DNA comprising a multiple STOP codon cassette. The remaining steps were the same as for the generation of *Tet2* His-HA knock-in mESCs. In addition to sequencing, the phenotype of the presumptive KO clones was confirmed by WB using an anti-TET2 antibody (Abcam ab124297, see Extended Data Fig. 4C).

Plasmids and sequences

Guide RNAs were cloned into pSpCas9n(BB)-2A-Puro (PX462) V2.0 vector (Addgene plasmid # 62987). All oligonucleotide and synthetic DNA sequences used are in Supplementary Table 4.

Protein immunoprecipitation

Cells were lysed in lysis buffer (10 mM HEPES pH 8.0$_{4\ °C}$, 350 mM KCl, 0.5% IGEPAL CA-630) followed by sonication on ice. Cell lysate was centrifuged at 18,000 g for 5 minutes and supernatant was taken. Cell extracts were incubated with HA antibody (Abcam ab9110) or IgG for 3 h at 4 °C and target proteins were recovered with protein G beads. Beads were washed with lysis buffer twice. Proteins were eluted from the beads by boiling in LDS loading buffer (Thermo Fisher) and resolved on 8% bis-tris gels. After being transferred to nitrocellulose membrane, signal was imaged.

CLIP

Tagged mESCs were pulsed with 500 μM 4SU for 4 hours, crosslinked with 400 mJ/cm^2 UVB (312 nm), and lysed in lysis buffer (10 mM HEPES pH 8.0, 350 mM KCl, 0.5% IGEPAL CA-630) with protease and RNase inhibitors. A sonication step was used to

increase lysis efficiency. His- and HA-fused proteins were first bound to Dynabeads His-Tag Isolation and Pulldown (Thermo Fisher) in lysis buffer for 1 h at 4°C. Beads were washed once using lysis buffer, twice using urea wash buffer (10 mM HEPES pH 8.0, 350 mM KCl, 0.5% IGEPAL CA-630, 8 M urea) and once using wash buffer (20 mM Tris pH 7.4$_{RT}$, 125 mM KCl, 800 mM imidazole, adjusted to pH 8 with KOH). Proteins were eluted by heating beads with 30 μL SDS elution buffer (20 mM Tris pH 7.4$_{RT}$, 5 mM EDTA, 125 mM NaCl, 2% SDS) at 70 °C for 10 minutes and diluted in 1.8 mL dilution buffer (20 mM Tris pH 7.4$_{RT}$, 5 mM EDTA, 125 mM NaCl) with protease inhibitors and RNase inhibitor. Next, proteins were incubated with HA antibody (Abcam ab9110) for 1 h at 4 °C and recovered with protein G Dynabeads by incubating at 4 °C for 45 minutes. DNA was removed with DNase, crosslinked RNA was dephosphorylated with FastAP enzyme and T4 PNK, and a fluorescently labelled RNA adapter was ligated to the 3' (Supplementary Table 4). Labeled complexes were eluted using 1x LDS loading buffer (Thermo Fisher) and resolved on 8% bis-tris gels, transferred to nitrocellulose membrane, and imaged.

CLIP-seq library construction

To generate size-matched input libraries, 0.2% of the cell lysate was treated with DNase and RNAs were partially digested with RNase. Input samples were loaded together with immunoprecipitated samples onto 8% bis-tris gels. A region ~75 kD above protein size was cut from the membrane and RNA was isolated by protease K treatment. All the remaining steps were performed according to eCLIP procedure[36] but using a redesigned 3' adapter labeled with an IR800 fluorochrome, similar to the iCLIP strategy[61], and a redesigned 5' adapter for cDNA ligation containing a 8-nt UMI (Supplementary Table 4). Libraries were sequenced on an Illumina NextSeq 500.

CLIP-seq analysis

Adapters were removed from reads with the AdapterRemoval tool[62] and reads smaller than 26 bp, including the 8 bp unique molecular identifier (UMI) were discarded. All the mapped reads were deduplicated using UMI-tools[63] based on UMI barcode information. Reads were then mapped against the mouse genome (mm10) with STAR (v 2.5.3a). To analyze data at the gene level (for Fig. 2A–C), reads were assigned to gene models using the R package DEGseq[64]. Peak calling was done according to a previous publication[39] with minor modifications. Briefly, the number of reads for both CLIP and input samples that mapped to each genomic region recovered as well as the total reads in each library, were compared using one-sided Fisher's exact tests. Regions where the test resulted in an FDR-adjusted P-value $< 10^{-5}$ were defined as peaks.

RIP-seq for hm5C

Total RNA was isolated from ~100 million cells using TriPure Isolation Reagent (Roche) and dissolved in BTE (10 mM bis-tris pH 6.7, 1 mM EDTA). Total RNA (1 mg) was diluted in 1 mL BTE heated at 70 °C for 10 minutes and transferred rapidly to ice to remove secondary structure. An equal volume of 2x immunoprecipitation buffer was added (40 mM Tris pH 7.4$_{RT}$, 1 mM EDTA pH 8.0, 700 mM NaCl, 0.2% NP-40) and the RNA was incubated with 12.5 μg of hm5C antibody (Active Motif #39769[21]) at 4 °C overnight. Immunoprecipitated RNA was recovered using protein G Dynabeads. After washing three

times with wash buffer (20 mM Tris-HCl pH 7.4$_{RT}$, 0.5 mM EDTA, 350 mM NaCl, 0.1% NP-40), RNA was purified from beads using TriPure Isolation Reagent. Sequencing libraries were prepared using the Illumina TruSeq small RNA kit and sequenced on an Illumina Nextseq 500.

Analysis of hm5C RIP-seq

Raw sequencing data was processed as for CLIP-seq but without the UMI-based deduplication step. For adapter removal the minimal length was set to 18. For peak calling, small RNA-seq reads were used as input samples. After Fisher's exact tests, P-values were FDR-corrected for multiple testing using Benjamini–Hochberg. IP regions with FDR-adjusted P-value $< 10^{-5}$ were defined as peaks as above.

RNA mass spectrometry

Total RNA was extracted from WT, *Tet2* KO, and *Tet1/2/3* tKO with TRIzol. The small RNA fraction (< 200 nts) was purified using RNA clean & concentrator columns (Zymo Research). For tRNA isolation, total RNA was resolved on a 6% polyacrylamide, 7 M urea gel, and the band corresponding to tRNAs (~70 nts) was excised and eluted overnight in elution buffer (10 mM bis-tris pH 6.7, 300 mM NaCl, 1 mM EDTA) at 4 ˚C. Eluted material was purified by acid phenol-chloroform extraction and ethanol precipitation. In some cases, heavy labeled standards of cytidine ($^{13}C_9, ^{15}N_3$ 5'-triphosphate, Sigma-Aldrich) and hm5C ($^{13}C, D_2$, Toronto Research Chemicals) were spiked into samples before digestion.

RNA samples ranging from 0.5 to 2 μg were digested to nucleosides in reaction buffer (1 mM ZnCl$_2$, 30 mM NAOAc pH 7.5) by 5 mU/μL of nuclease P1, 6.25 μU/μL of phosphodiesterase II, 5 mU/μL of recombinant shrimp alkaline phosphatase, and 500 μUnits/μL of phosphodiesterase I overnight at room temperature. Samples were purified by using in-house stop-and-go-extraction tips (StageTips) topped with Thermo Hypercarb porous graphic carbon. StageTips were conditioned with acetonitrile (ACN) and washed with 0.1% formic acid (FA). Samples were loaded, washed with 0.1% FA, and then eluted in 70% acetonitrile. Samples were dried in a Savant SpeedVac and submitted for LC-MS/MS analysis. LC was performed on a Thermo Vanquish Flex Binary UPLC with a Thermo Accucore Vanquish C18 column (150 × 2.1 mm, 1.5 μm) at 60 ˚C using 0.1% FA as buffer A and 0.1% FA in ACN as buffer B. Nucleosides were separated with a gradient of 0% B to 2% B over 7 minutes. Mass spectra were acquired for C by fragmenting 244.093 m/z (11 V collision energy, 2 mTorr CID gas) and detecting 112.050 m/z. Mass spectra were acquired for hm5C by fragmenting 274.103 m/z (11 V collision energy, 2 mTorr CID gas) and detecting 142.061 m/z. Data analysis was performed by hand.

Transient transfections

ES cells were transfected using Lipofectamine 3000 reagent following the manufacturer's guidelines with the following modifications. To achieve comparable levels of protein expression the quantity of protein-coding vector was adjusted across treatments with 6.6 μg, 13 μg, and 20 μg of DNA transfected of the GFP, TET2 WT, and TET2 HxD mutant plasmids respectively. Empty vector was used to bring the total amount of DNA to 20 μg for each condition, repeated in triplicate. The DNA-Lipofectamine complexes were added to 2.5

million cells in suspension and incubated for 15 minutes at room temperature. Ten percent of the cell suspension was transferred into a 6-well dish to confirm protein expression; the remaining cells were transferred to a gelatin-coated 10 cm plate. Cells were maintained for 42 hours and RNA was harvested with TRIzol.

In vitro TET2 reaction

The ability of recombinant TET2 to convert m5C to hm5C on tRNAs *in vitro* was assayed as described in Liu *et al.*[65] and DeNizio *et al.*[20]. Tranfer RNAs from *Tet1/2/3* tKO were purified by size selection on a polyacrylamide gel as described above and then used as substrate in the *in vitro* reaction. Purified tRNAs were incubated at a concentration of 2 μM in 50 mM HEPES pH 7.5, 100 mM NaCl, 1 mM DTT, 2 mM ascorbic acid, 1 mM alpha-keto-glutarate, 75 μM ammonium iron(II) sultfate with or without 5 pmol recombinant mouse TET2 catalytic domain fragment (residues 1,042–1,921[65]). The reaction was assembled on ice and then incubated at 37 ˚C for 1 h. At the end of the reaction tRNAs were analyzed by mass spectrometry as described above.

Small RNA sequencing

Small RNA sequencing libraries were prepared from RNA extracted from WT and *Tet2*[−/−] (Fig. 5) or from E14 mESCs and *Tet2*[−/−] (Extended Data Fig. 5). Total RNA was extracted using TRIzol (Thermo Fisher Scientific) and 80% EtOH washes were performed during precipitation. For small RNA cloning, total RNA was size selected for 14–38 nts on 15% Novex TBE urea gels (Thermo Fisher Scientific). Small RNA libraries were prepared using the Illumina TruSeq small RNA kit (Fig. 5) or the NEB small RNA-seq library kit (Extended Data Fig. 5) and sequenced on a NextSeq 500 platform.

Small RNA-seq analysis

Small RNA reads were quality filtered using Gordon Assaf's fastx-toolkit. AdapterRemoval was used to clip Illumina adapters and remove any Truseq Illumina stop oligo sequences. Reads were aligned to the mouse mm10 UCSC genome annotation using Bowtie2 to assign multi-mapping reads in an unbiased, random way. Aligned reads were filtered for 0–2 mismatches using Samtools and Bamtools. For tRF analysis, reads were sorted into non-CCA and CCA-ending sequences (CCA clipped), and aligned against the UCSC tRNA gene annotation. CCA reads derived from tRNAs aligned to the 3' end of the mature tRNA; the terminal CCA was assigned as the zero position in tRNA coverage plots. Non-CCA tRNA reads were plotted along tRNA coordinates defining the 5' end as the zero position. We defined tRF5 as non-CCA reads 28–35 nts long overlapping the 5' end of the mature tRNA sequence ± 5 nts to allow for imprecision in the genomic annotation. tRNA fragments ending in CCA that were 17–19 nts and 22 nts long were named tRF3a and tRF3b, respectively, according to the nomenclature of the Dutta laboratory[66].

All read counts were normalized to total aligned reads per library including the CCA reads that were aligned separately, resulting in reads per million mapped (RPMs). All reads assigned to gene models, including tRNA gene models, were imported into the R package DESeq2[67] for differential expression analyses. For overlap analyses, we only considered tRNAs that were detected in at least one sample of the two datasets being overlapped and we

collapsed tRNAs and tRFs with identical sequences to minimize overlap inflation. To calculate *P*-values from the hypergeometric distribution we used a conservative "universe" number that only included detected tRNAs.

Reanalysis of BS-seq

The tRNA BS-seq data from Legrand *et al*[10] were downloaded from the GEO (GSE81825). tRNA annotations were lifted over from mm9 to mm10. The positions of the m5C sites were obtained directly from the tables obtained from the GEO and only sites with coverage (reads spanning the site) of at least 10 were considered.

Statistics

Sample size and statistical tests are indicated in the figure legends when necessary. Unless otherwise noted all statistical tests were two-sided. All replicates were obtained by measuring distinct samples (biological and/or experimental replicates) and not by measuring multiple times the same sample (technical replicates). Boxplots (Fig. 5 and Extended Data Fig. 5) were drawn using default parameters in R (center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range).

Reporting Summary statement

Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Code availability

Software utilized for each analysis is detailed in the relevant method section. Scripts and R markdowns documents to generate figures are available from the corresponding author upon request.

Data availability

RNA sequencing data generated for this study have been deposited in the NCBI GEO with accession number GSE133472. Raw mass spectrometry data are available on figshare (doi: 10.6084/m9.figshare.c.5133581).
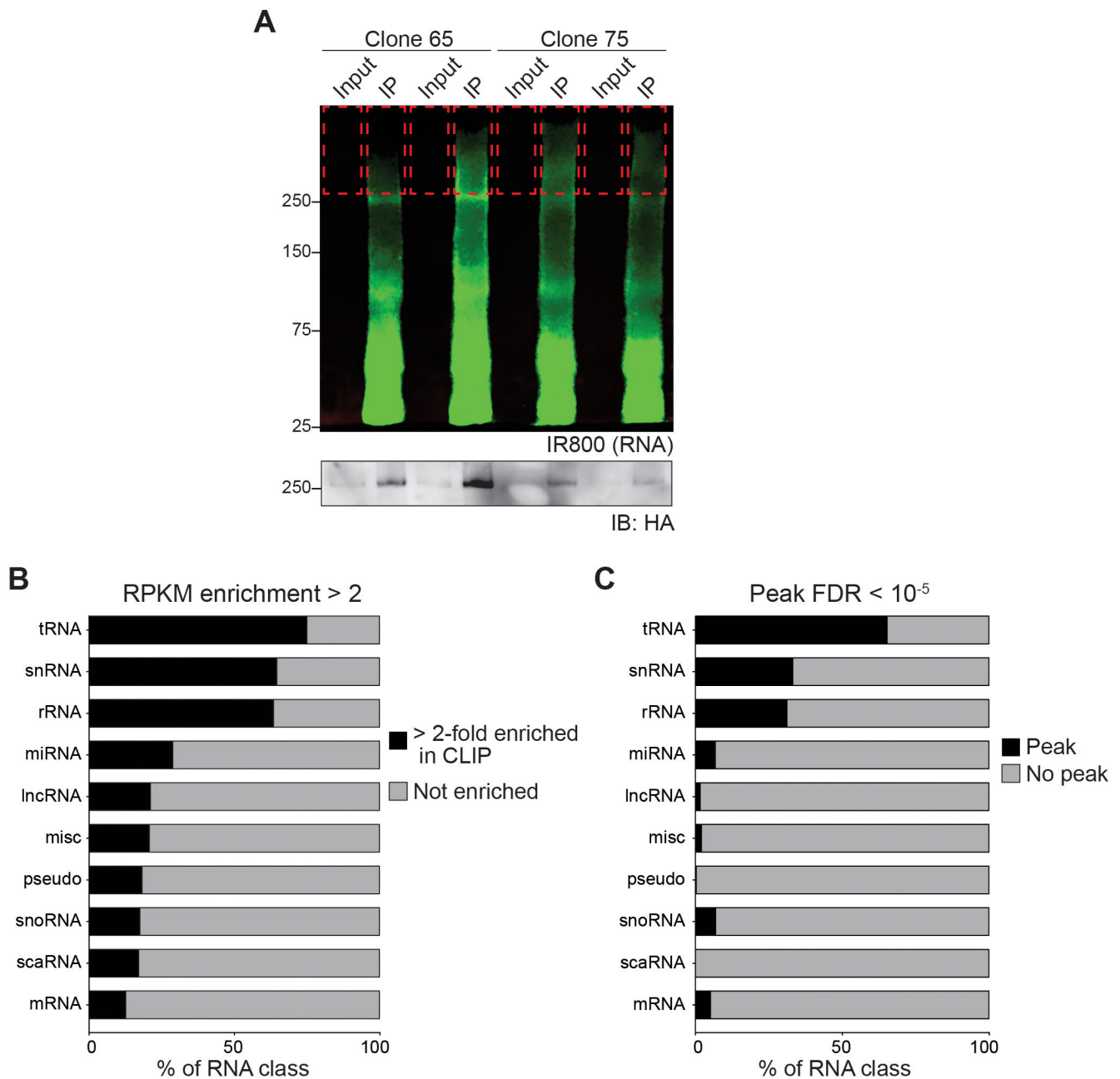
Extended Data

A

B
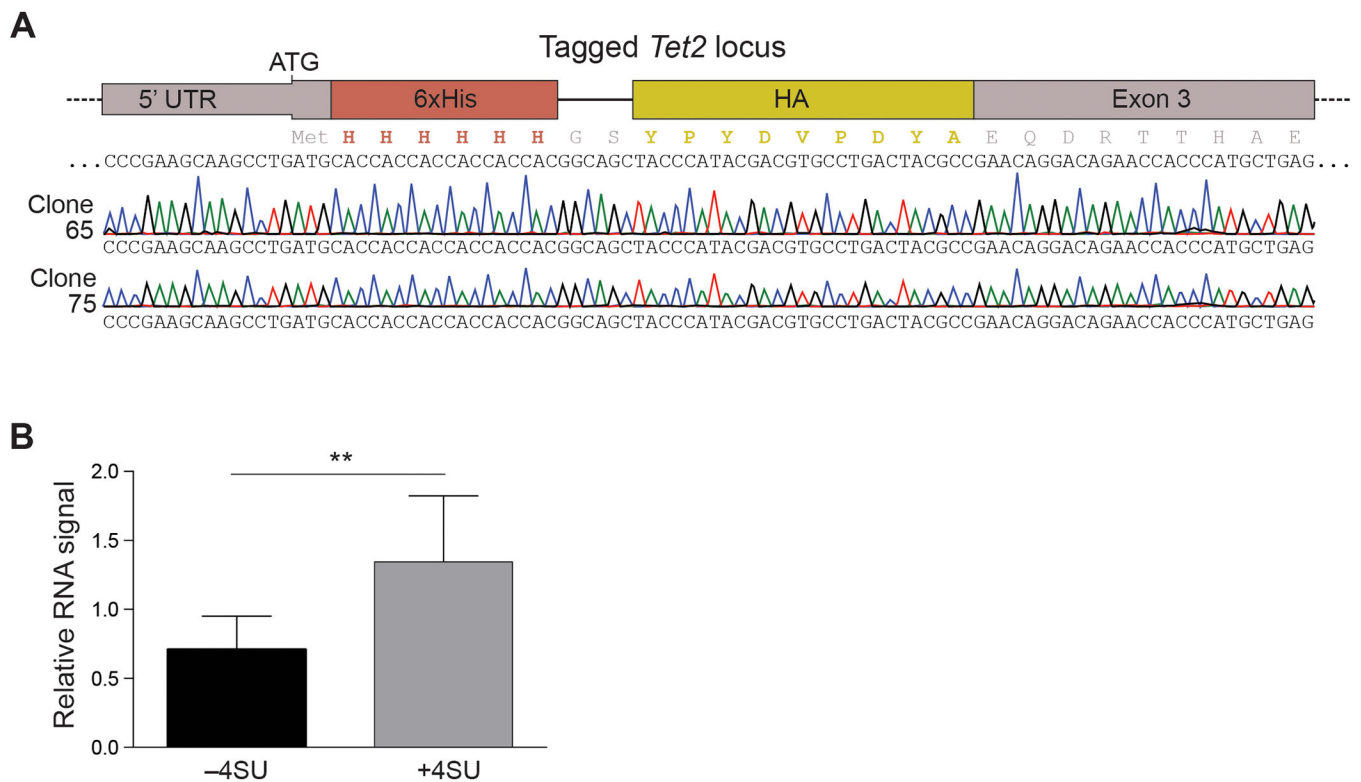
Extended Data Fig. 1. Generation of epitope-tagged *Tet2* alleles and CLIP quantification

A) Genotype validation for 6xHis–HA knock-in at the *Tet2* locus by Sanger sequencing. The targeted allele scheme (top), expected protein and DNA sequence (middle), and sequencing traces (bottom) for the two clones used in subsequent experiments are shown.

(B) Quantification of Fig. 1D; fluorescence signal (crosslinked RNA) was normalized to WB signal (protein). Bars represent the mean + s.e.m. *P*-value is from a Student's *t*-test.

He et al.

Page 16

Extended Data Fig. 2. Additional analyses on TET2 CLIP-seq

(A) Fluorescence image of two CLIP replicates from two cell lines (#65 and #75) used for CLIP-seq library construction. The dashed red boxes indicate the position of the excised bands. Bottom panel, Western blot for HA was used as a loading control (bottom). Uncropped blot images are shown in Supplementary Fig. 1.

(B) Percentage of transcripts from the indicated classes enriched > 2-fold in TET2 CLIP compared to input (black bars). The non-enriched portion of each class is shown in gray. Only transcripts detected (> 1 read) in at least one replicate were considered. snRNA, small nuclear RNAs; rRNA, ribosomal RNAs; miRNA, micro RNAs; lncRNA, long noncoding RNAs; misc, RNAs not included in the other displayed categories; pseudo, pseudogene-

*Nat Struct Mol Biol*. Author manuscript; available in PMC 2021 May 23.

derived RNAs; snoRNA, small nucleolar RNAs; scaRNA, small Cajal body RNAs; mRNA, protein-coding messenger RNAs.

(C) Same as (B) but considering as enriched only RNAs containing a peak (see text for details) with FDR-corrected P-value $< 10^{-5}$.
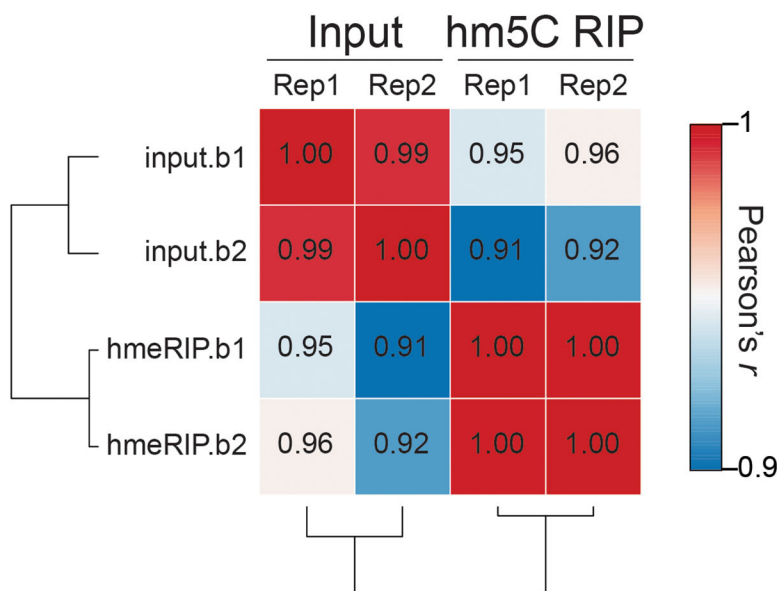
Extended Data Fig. 3. Replicate consistency for the hm5C RIP experiment
Clustered heatmap showing the Pearson's correlation coefficients between RPKMs calculated on all annotated genes in input and hm5C RIP for two biological replicates each.

A

SYBR gold

B

C

D
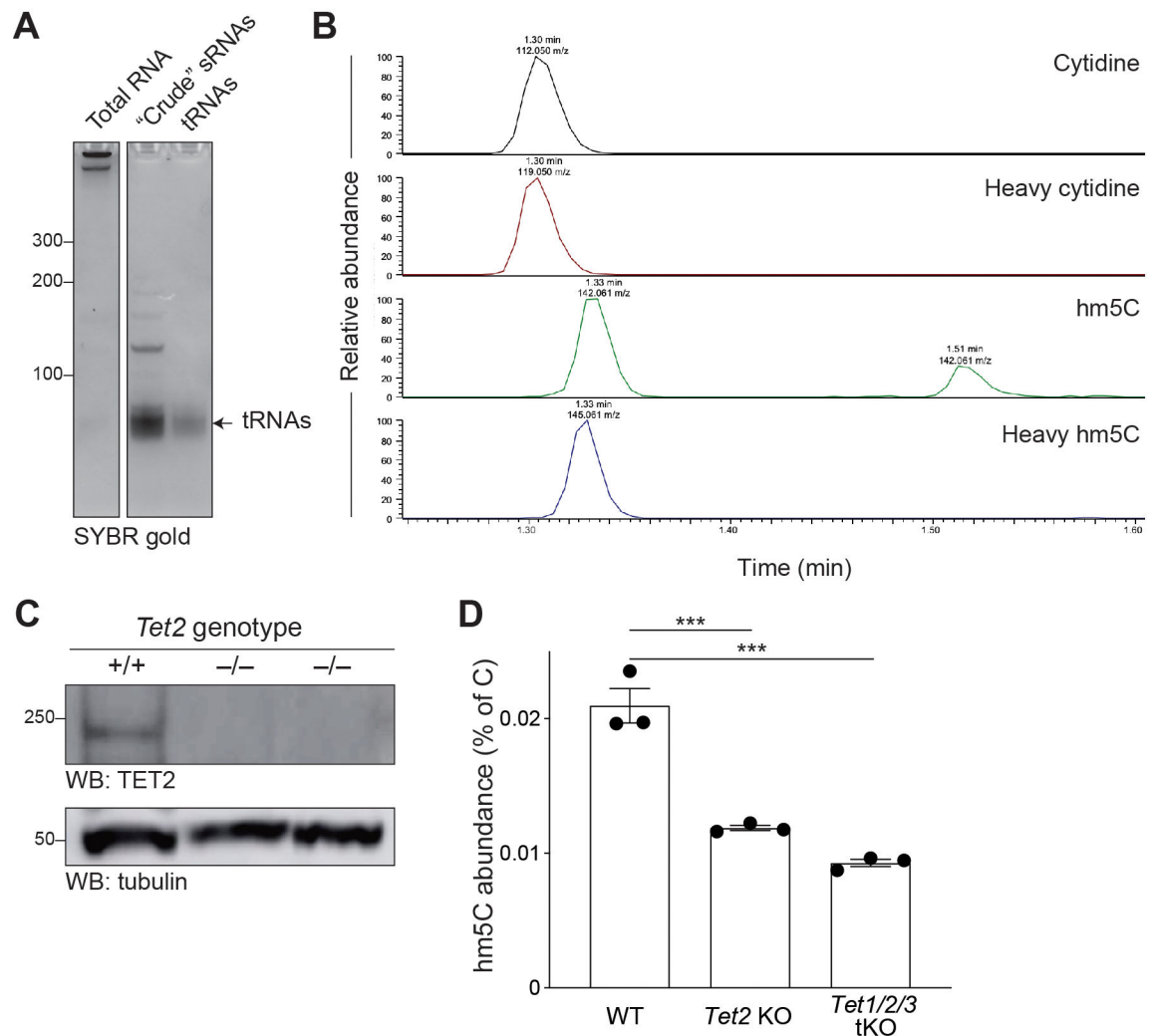
**Extended Data Fig. 4. RNA fractionation and mass spectrometry**

(A) Representative gel (6% polyacrylamide, 7 M urea) showing the three RNA fractions analyzed in Fig. 4: total RNA (no fractionation), small RNAs < 200 nts (column-based size selection), and tRNAs (gel purification). The band corresponding to tRNAs (~70 nts) is indicated by the arrow.

(B) Mass spectrometry chromatograms of nucleosides fragmented into nucleobases. Data were acquired by isolating a precursor ion (nucleoside), fragmenting the precursor ion, and then isolating and detecting a known fragment ion (nucleobase). The top two chromatograms show C (244.093 → 112.050 m/z) and a spiked-in heavy C standard. The bottom two chromatograms show hm5C (256.103 → 142.061 m/z) and a spiked-in heavy hm5C standard (277.103 → 145.061 m/z). The representative chromatograms shown were obtained from the same run on tRNAs from *Tet1/2/3 tKO* cells transiently transfected with TET2 WT (Fig. 4D). Only one known nucleoside, 5-aminomethyluridine (nm5U), is isobaric with hm5C, and can be easily distinguished from hm5C by retention time.

(C) Western blot for TET2 comparing WT (+/+) and presumptive KO (–/–) clones as determined by PCR screening and Sanger sequencing. Tubulin is shown as loading control.

(D) Mass spectrometry quantification of hm5C in size-selected small RNAs < 200 nts from WT (left) *Tet2* single KO (middle) and *Tet1/2/3* triple KO (right) mESCs. Bars represent mean + s.e.m. ***, $P < 0.001$. *P*-values are from one-way ANOVA followed by Holm-Sidak test.

Uncropped gel and blot images for A and C are shown in Supplementary Fig. 1.
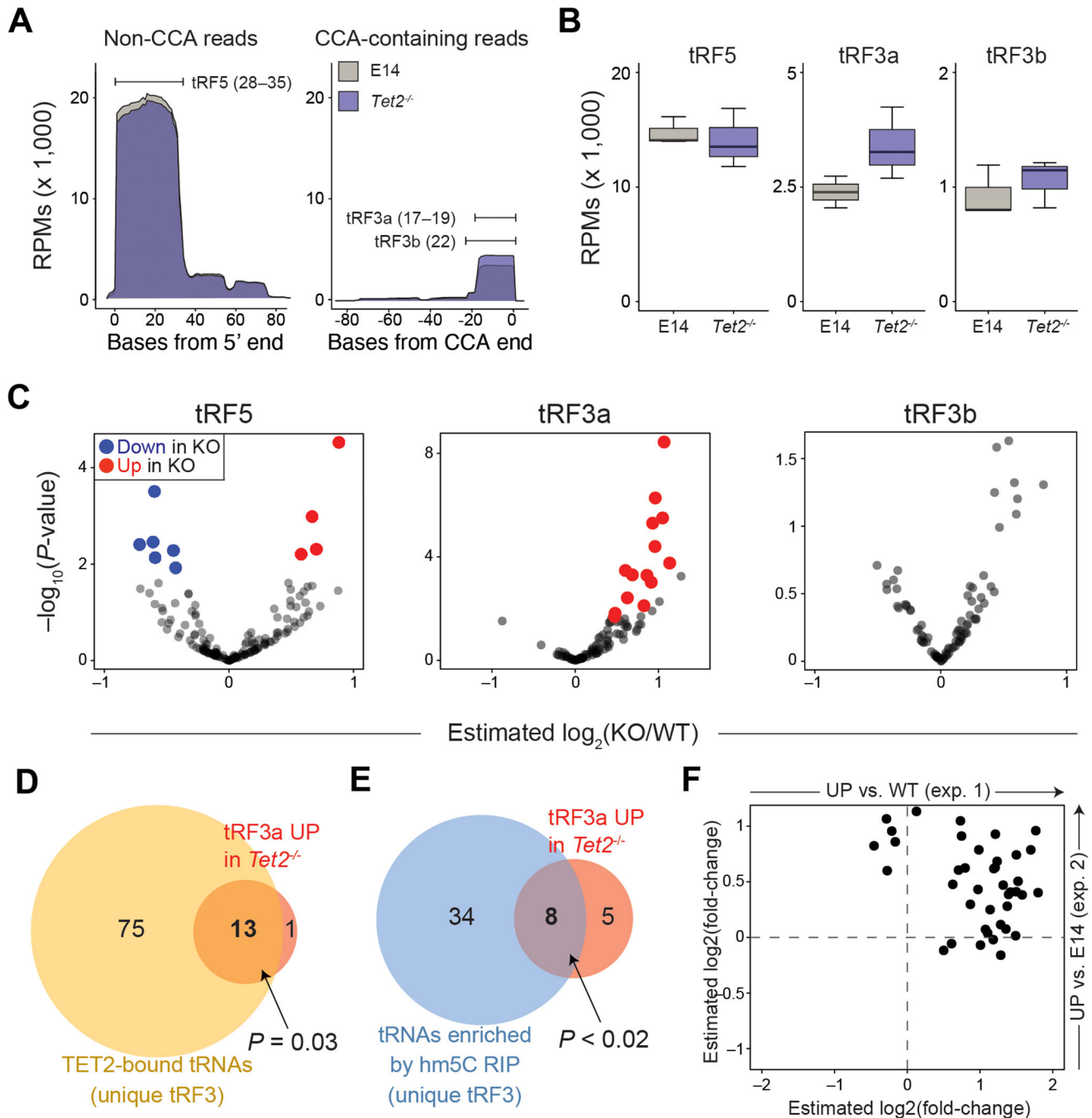
Extended Data Fig. 5. Additional small RNA sequencing dataset comparing *Tet2*[−/−] with E14 ESCs

A) Coverage of tRNA genes by non-CCA (left) and CCA-containing (right) reads in small RNAs purified from control (E14) or $Tet2^{-/-}$ cells. Plots show the average RPMs. Position of the three types of tRFs discussed in the text are indidated (tRF5, tRF3a, and tRF3b).

(B) Quantification of (A) but only considering size-filtered reads; 28–35 for tRF5, 17–19 (inclusive of CCA) for tRF3a, and 22 (inclusive of CCA) for tRF3b.

(C) Differential expression analysis for individual tRFs in E14 and $Tet2^{-/-}$ cells. Estimated (DESeq2) fold changes are plotted on the x axis and the log-converted P-value on the y axis. Blue and red dots highlight individual tRFs that pass an adjusted P-value cutoff of 0.1 and are downregulated or upregulated in the KO, respectively.

(D) Overlap of TET2-bound tRNAs as determined by CLIP (Fig. 2), and the tRF3a significantly upregulated in $Tet2^{-/-}$ cells as determined in (C). The TET2-bound tRNAs were grouped according to the predicted sequence of the tRF3 produced from them. The P-value was calculated based on the hypergeometric distribution.

(E) Same as (D) but showing the overlap with tRNAs enriched by hm5C RIP (Fig. 3).

(F) Comparison of estimated $\log_2$(fold-changes) for all tRF3a significantly enriched in $Tet2^{-/-}$ cells in two independent experiments (exp 1, Fig. 5; and exp 2, Extended Data Fig. 5).

Replicates are from three independent cell cultures and RNA purifications per genotype.
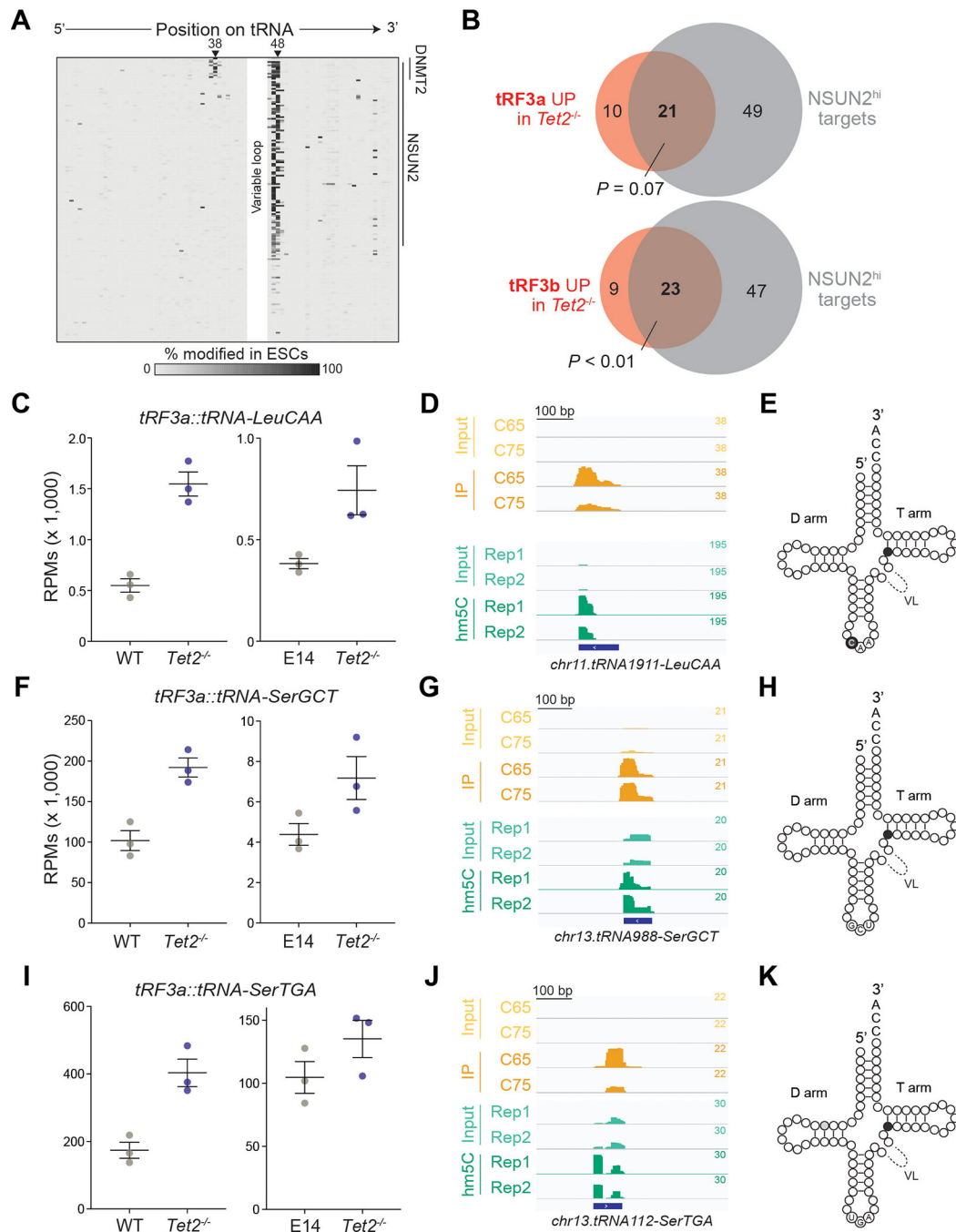
Extended Data Fig. 6. Examples of tRNAs methylated by NSUN2 and regulated by TET2

(A) Heatmap for the % of unconverted BS-seq reads on tRNAs as reported by Legrand *et al*[10] (GEO series GSE81825).

(B) Overlap of tRF3a (top) or tRF3b (bottom) detected as significantly upregulated in *Tet2*[−/−] cells compared to WT in Fig. 5 with highly methylated targets of NSUN2 (> 75% m5C at NSUN2 sites). *P*-values are from Fisher's test comparing overlaps with methylated and unmethylated tRNAs.

(C) Levels (RPMs) for a tRF3a from LeuCAA tRNAs in WT and *Tet2*$^{-/-}$ cells. The two plots show data from two independent experimental replicates, corresponding to Fig. 5 (left) and Extended Fig. 5 (right). Mean ± s.e.m. are shown.

(D) Genomic browser snapshot for CLIP and hm5C at the *chr11.tRNA1911-LeuCAA* locus. Matching inputs are shown. The *y* axis represents RPMs.

(E) Schematic depiction of methylation patterns on *chr11.tRNA1911-LeuCAA* as determined by BS-seq in Legrand *et al*[10]. The position of m5C in the anticodon and after the variable loop (VL) is indicated by thicker circles and the % of uncovered reads is shown using the same color scale used in (A).

(F–H) Same as (C–E) but for *chr13.tRNA988-SerGCT*.

(I–K) Same as (C–E) but for *chr13.tRNA112-SerTGA*.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

REFERENCES

1. Roundtree IA, Evans ME, Pan T & He C Dynamic RNA Modifications in Gene Expression Regulation. Cell 169, 1187–1200 (2017). [PubMed: 28622506]

2. Gilbert WV, Bell TA & Schaening C Messenger RNA modifications: Form, distribution, and function. Science 352, 1408–12 (2016). [PubMed: 27313037]

3. Alarcon CR, Lee H, Goodarzi H, Halberg N & Tavazoie SF N6-methyladenosine marks primary microRNAs for processing. Nature 519, 482–5 (2015). [PubMed: 25799998]

4. Patil DP et al. m(6)A RNA methylation promotes XIST-mediated transcriptional repression. Nature 537, 369–373 (2016). [PubMed: 27602518]

5. Ke S et al. m(6)A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. Genes Dev 31, 990–1006 (2017). [PubMed: 28637692]

6. Squires JE et al. Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res 40, 5023–33 (2012). [PubMed: 22344696]

7. Delatte B et al. RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. Science 351, 282–5 (2016). [PubMed: 26816380]

8. Huang T, Chen W, Liu J, Gu N & Zhang R Genome-wide identification of mRNA 5-methylcytosine in mammals. Nat Struct Mol Biol 26, 380–388 (2019). [PubMed: 31061524]

9. Hussain S, Aleksic J, Blanco S, Dietmann S & Frye M Characterizing 5-methylcytosine in the mammalian epitranscriptome. Genome Biol 14, 215 (2013). [PubMed: 24286375]

10. Legrand C et al. Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. Genome Res 27, 1589–1596 (2017). [PubMed: 28684555]

11. Schaefer M et al. RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. Genes Dev 24, 1590–5 (2010). [PubMed: 20679393]

12. Tuorto F et al. RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. Nat Struct Mol Biol 19, 900–5 (2012). [PubMed: 22885326]

13. Gkatza NA et al. Cytosine-5 RNA methylation links protein synthesis to cell metabolism. PLoS Biol 17, e3000297 (2019). [PubMed: 31199786]

14. Blanco S et al. Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders. EMBO J 33, 2020–39 (2014). [PubMed: 25063673]

15. Wu X & Zhang Y TET-mediated active DNA demethylation: mechanism, function and beyond. Nat Rev Genet 18, 517–534 (2017). [PubMed: 28555658]

16. Dawlaty MM et al. Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. Dev Cell 29, 102–11 (2014). [PubMed: 24735881]

17. Li X et al. Ten-eleven translocation 2 interacts with forkhead box O3 and regulates adult neurogenesis. Nat Commun 8, 15903 (2017). [PubMed: 28660881]

18. Shen Q et al. Tet2 promotes pathogen infection-induced myelopoiesis through mRNA oxidation. Nature 554, 123–127 (2018). [PubMed: 29364877]

19. Fu L et al. Tet-mediated formation of 5-hydroxymethylcytosine in RNA. J Am Chem Soc 136, 11582–5 (2014). [PubMed: 25073028]

20. DeNizio JE, Liu MY, Leddin EM, Cisneros GA & Kohli RM Selectivity and Promiscuity in TET-Mediated Oxidation of 5-Methylcytosine in DNA and RNA. Biochemistry 58, 411–421 (2019). [PubMed: 30387995]

21. Guallar D et al. RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells. Nat Genet 50, 443–451 (2018). [PubMed: 29483655]

22. Kumar P, Kuscu C & Dutta A Biogenesis and Function of Transfer RNA-Related Fragments (tRFs). Trends Biochem Sci 41, 679–689 (2016). [PubMed: 27263052]

23. Ivanov P, Emara MM, Villen J, Gygi SP & Anderson P Angiogenin-induced tRNA fragments inhibit translation initiation. Mol Cell 43, 613–23 (2011). [PubMed: 21855800]

24. Sobala A & Hutvagner G Small RNAs derived from the 5' end of tRNA can inhibit protein translation in human cells. RNA Biol 10, 553–63 (2013). [PubMed: 23563448]

25. Blanco S et al. Stem cell function and stress response are controlled by protein synthesis. Nature 534, 335–40 (2016). [PubMed: 27306184]

26. Schorn AJ, Gutbrod MJ, LeBlanc C & Martienssen R LTR-Retrotransposon Control by tRNA-Derived Small RNAs. Cell 170, 61–71 e11 (2017). [PubMed: 28666125]

27. Chen Q et al. Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. Science 351, 397–400 (2016). [PubMed: 26721680]

28. Sharma U et al. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. Science 351, 391–396 (2016). [PubMed: 26721685]

29. Conine CC, Sun F, Song L, Rivera-Perez JA & Rando OJ Small RNAs Gained during Epididymal Transit of Sperm Are Essential for Embryonic Development in Mice. Dev Cell 46, 470–480 e3 (2018). [PubMed: 30057276]

30. Sharma U et al. Small RNAs Are Trafficked from the Epididymis to Developing Mammalian Sperm. Dev Cell 46, 481–494 e6 (2018). [PubMed: 30057273]

31. He C et al. High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. Mol Cell 64, 416–430 (2016). [PubMed: 27768875]

32. Koh KP et al. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell 8, 200–13 (2011). [PubMed: 21295276]

33. Pastor WA, Aravind L & Rao A TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nat Rev Mol Cell Biol 14, 341–56 (2013). [PubMed: 23698584]

34. Hafner M et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141, 129–41 (2010). [PubMed: 20371350]

35. Friedersdorf MB & Keene JD Advancing the functional utility of PAR-CLIP by quantifying background binding to mRNAs and lncRNAs. Genome Biol 15, R2 (2014). [PubMed: 24393468]

36. Van Nostrand EL et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat Methods 13, 508–14 (2016). [PubMed: 27018577]

37. Gajjar M et al. The p53 mRNA-Mdm2 interaction controls Mdm2 nuclear trafficking and is required for p53 activation following DNA damage. Cancer Cell 21, 25–35 (2012). [PubMed: 22264786]

38. Zarnegar BJ et al. irCLIP platform for efficient characterization of protein-RNA interactions. Nat Methods 13, 489–92 (2016). [PubMed: 27111506]

39. Meyer KD et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell 149, 1635–46 (2012). [PubMed: 22608085]

40. Hu X et al. Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. Cell Stem Cell 14, 512–22 (2014). [PubMed: 24529596]

41. Ko M et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature 468, 839–43 (2010). [PubMed: 21057493]

42. Huber SM, van Delft P, Tanpure A, Miska EA & Balasubramanian S 2'-O-Methyl-5-hydroxymethylcytidine: A Second Oxidative Derivative of 5-Methylcytidine in RNA. J Am Chem Soc 139, 1766–1769 (2017). [PubMed: 28107630]

43. Frye M & Blanco S Post-transcriptional modifications in development and stem cells. Development 143, 3871–3881 (2016). [PubMed: 27803056]

44. Schorn AJ & Martienssen R Tie-Break: Host and Retrotransposons Play tRNA. Trends Cell Biol 28, 793–806 (2018). [PubMed: 29934075]

45. Jin SG et al. Tet3 Reads 5-Carboxylcytosine through Its CXXC Domain and Is a Potential Guardian against Neurodegeneration. Cell Rep 14, 493–505 (2016). [PubMed: 26774490]

46. Beck DB et al. Delineation of a Human Mendelian Disorder of the DNA Demethylation Machinery: TET3 Deficiency. Am J Hum Genet (2020).

47. Dunn DB Additional components in ribonucleic acid of rat-liver fractions. Biochim Biophys Acta 34, 286–8 (1959). [PubMed: 13818674]

48. Klagsbrun M An evolutionary study of the methylation of transfer and ribosomal ribonucleic acid in prokaryote and eukaryote organisms. J Biol Chem 248, 2612–20 (1973). [PubMed: 4633356]

49. Schopman NC, Heynen S, Haasnoot J & Berkhout B A miRNA-tRNA mix-up: tRNA origin of proposed miRNA. RNA Biol 7, 573–6 (2010). [PubMed: 20818168]

50. Maute RL et al. tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. Proc Natl Acad Sci U S A 110, 1404–9 (2013). [PubMed: 23297232]

51. Li Z et al. Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. Nucleic Acids Res 40, 6787–99 (2012). [PubMed: 22492706]

52. Yamasaki S, Ivanov P, Hu GF & Anderson P Angiogenin cleaves tRNA and promotes stress-induced translational repression. J Cell Biol 185, 35–42 (2009). [PubMed: 19332886]

53. Genenncher B et al. Mutations in Cytosine-5 tRNA Methyltransferases Impact Mobile Element Expression and Genome Stability at Specific DNA Repeats. Cell Rep 22, 1861–1874 (2018). [PubMed: 29444437]

54. Goll MG et al. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science 311, 395–8 (2006). [PubMed: 16424344]

55. Hussain S et al. NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. Cell Rep 4, 255–61 (2013). [PubMed: 23871666]

56. Helm M Post-transcriptional nucleotide modification and alternative folding of RNA. Nucleic Acids Res 34, 721–33 (2006). [PubMed: 16452298]

57. Agris PF Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. EMBO Rep 9, 629–35 (2008). [PubMed: 18552770]

58. Anderson JT & Wang X Nuclear RNA surveillance: no sign of substrates tailing off. Crit Rev Biochem Mol Biol 44, 16–24 (2009). [PubMed: 19280429]

59. Motorin Y & Helm M tRNA stabilization by modified nucleotides. Biochemistry 49, 4934–44 (2010). [PubMed: 20459084]

60. Ying QL et al. The ground state of embryonic stem cell self-renewal. Nature 453, 519–23 (2008). [PubMed: 18497825]

61. Konig J et al. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat Struct Mol Biol 17, 909–15 (2010). [PubMed: 20601959]

62. Schubert M, Lindgreen S & Orlando L AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes 9, 88 (2016). [PubMed: 26868221]

63. Smith T, Heger A & Sudbery I UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res 27, 491–499 (2017). [PubMed: 28100584]

64. Wang L, Feng Z, Wang X, Wang X & Zhang X DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics 26, 136–8 (2010). [PubMed: 19855105]

65. Liu MY, DeNizio JE & Kohli RM Quantification of Oxidized 5-Methylcytosine Bases and TET Enzyme Activity. Methods Enzymol 573, 365–85 (2016). [PubMed: 27372762]

66. Kumar P, Anaya J, Mudunuri SB & Dutta A Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. BMC Biol 12, 78 (2014). [PubMed: 25270025]

67. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550 (2014). [PubMed: 25516281]
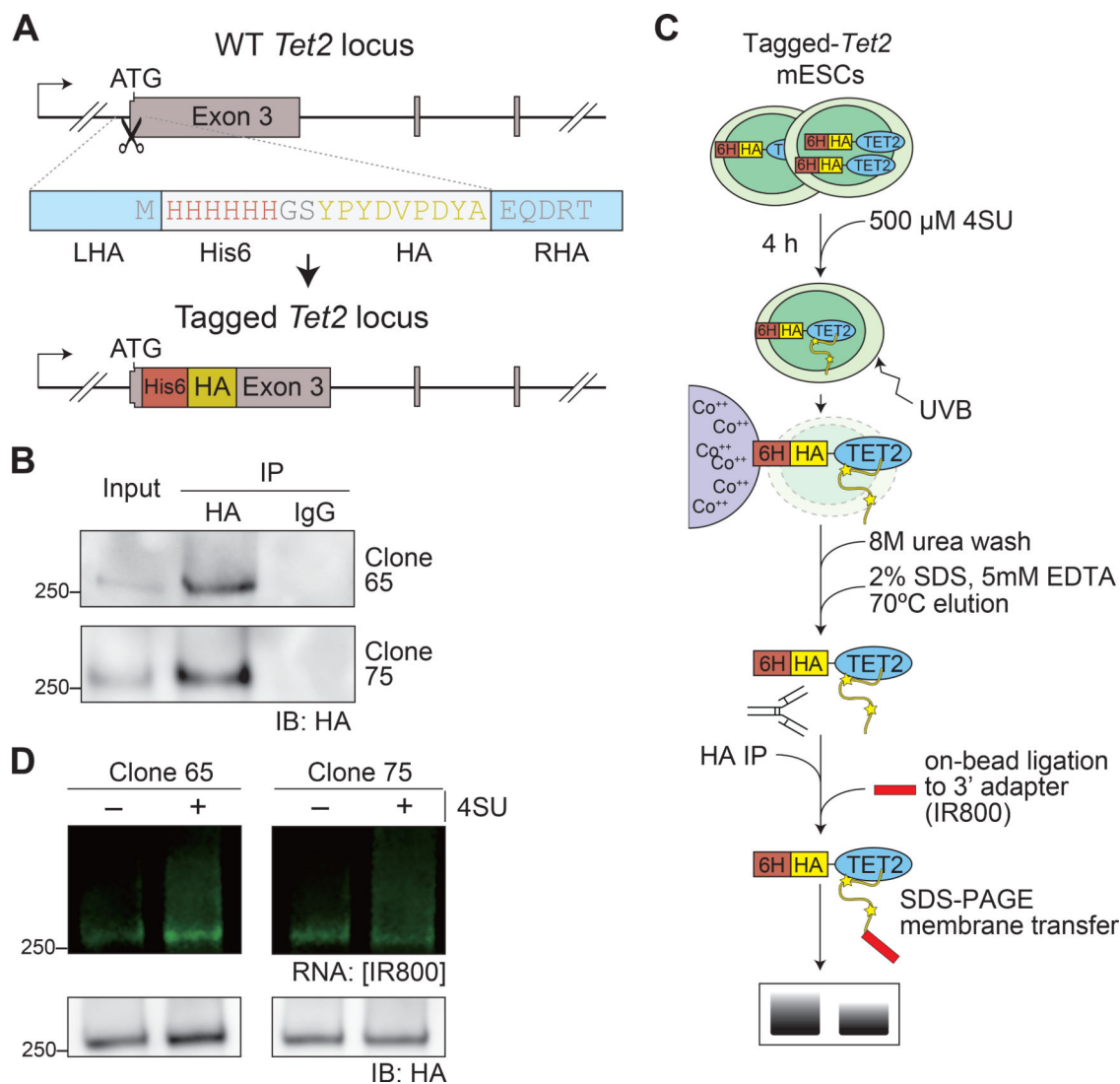
Figure 1. Validation of RNA binding by TET2 in mESCs

(A) Schematic depiction of the targeting strategy for the *Tet2* locus (NM_001040400.2). WT configuration (top), donor DNA (middle), and targeted allele (bottom) are shown. LHA, left homology arm; RHA, right homology arm.

(B) Immunoprecipitation of 6xHis–HA-tagged TET2 from genome-edited mESCs. Two positive clones (#65 and #75) were analyzed. IgG was used as a negative control.

(C) Experimental scheme for our CLIP protocol. Tagged mESCs were pulsed with 4SU and crosslinked with UVB (312 nm). 6xHis–HA-tagged TET2 was pulled down using cobalt-coated beads (step 1), washed in denaturing conditions, eluted, and immunoprecipitated (step 2) using an anti-HA antibody. Crosslinked RNAs were labelled by on-bead ligation to a fluorescent 3' RNA adapter. The crosslinked and immunoprecipitated ribonucleoprotein complexes were separated by SDS-PAGE and transferred to nitrocellulose for fluorescent imaging or for membrane elution and library construction.

(D) TET2 CLIP in mESCs expressing 6xHis–HA-TET2. The green fluorescence signal reports on the abundance of crosslinked RNA of different sizes (causing a delay in protein

migration that appears as a smear); mESCs not pulsed with 4SU (–4SU) were used as a negative control. The HA western blot (bottom panel) shows the amounts of TET2 protein present. Experiments were done in replicates using two separate 6xHis–HA-*Tet2* mESC clones.

Uncropped blot images for B and D are shown in Supplementary Figure 1.
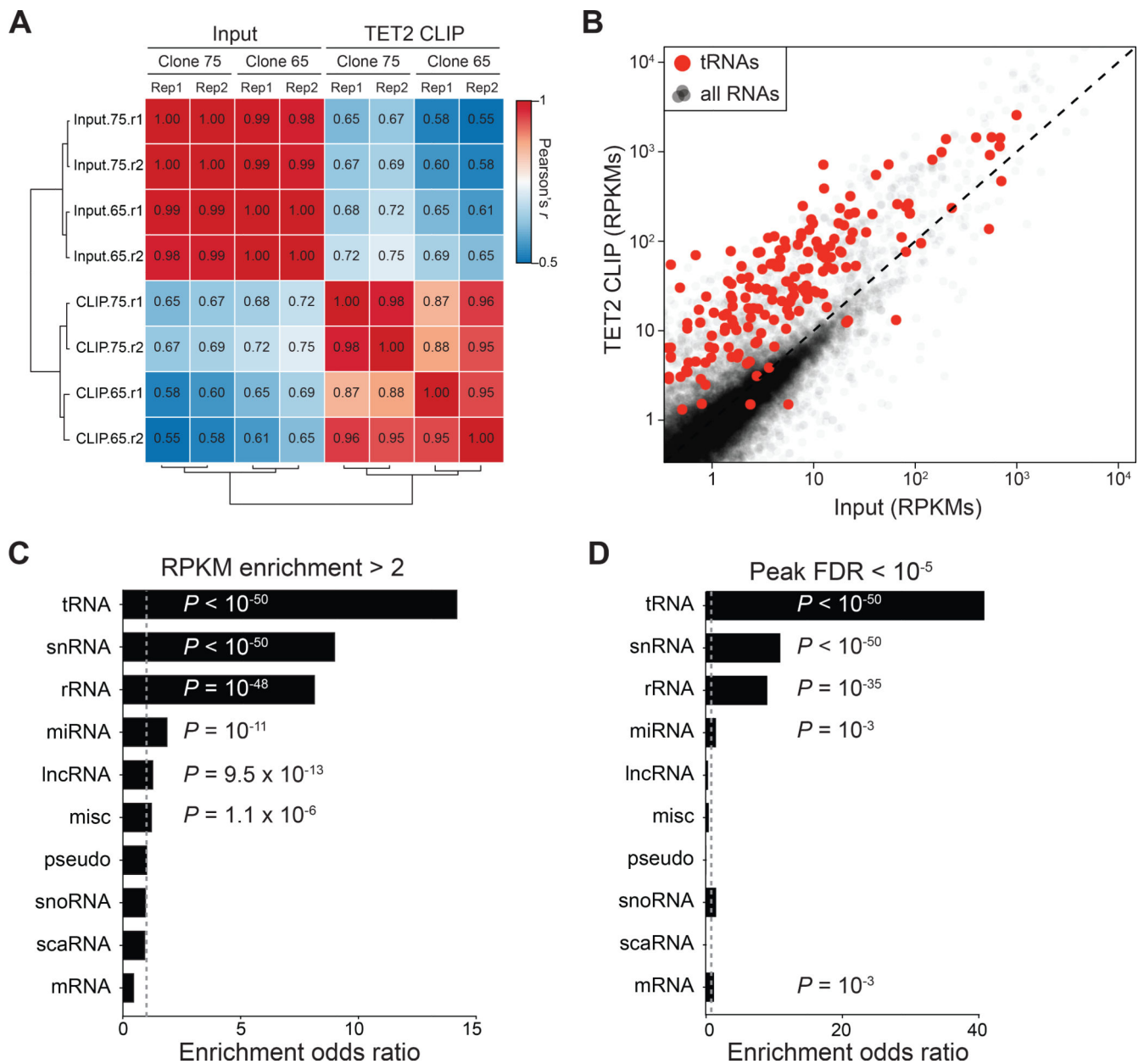
Figure 2. The TET2-bound transcriptome

(A) Clustered heatmap showing the Pearson's correlation coefficients between reads per kilobase per million (RPKMs) calculated on all annotated genes in input and TET2 CLIP from all replicates.

(B) Enrichment of transcripts in TET2 CLIP-seq. Black semi-transparent circles represent all detected RNAs. Transcripts annotated as tRNAs are highlighted in red. Mean RPKMs from input sample and CLIP are plotted on the *x* and *y* axis, respectively.

(C) Odds ratio for classes of > 2-fold enriched RNAs in TET2 CLIP. Fisher's one-sided test *P*-values are indicated, when significant (*P* < 0.05). Only transcripts detected (> 1 read) in at least one replicate were considered. snRNA, small nuclear RNAs; rRNA, ribosomal RNAs; miRNA, micro RNAs; lncRNA, long noncoding RNAs; misc, RNAs not included in the

other displayed categories; pseudo, pseudogene-derived RNAs; snoRNA, small nucleolar RNAs; scaRNA, small Cajal body RNAs; mRNA, protein-coding messenger RNAs.

(D) Same as (C) but considering only RNAs containing a peak (see text for details) with FDR-corrected P-value $< 10^{-5}$ as enriched.

Data are from four replicates: two independent immunoprecipitations from each of two independently generated epitope-tagged clones (clone 65 and 75).
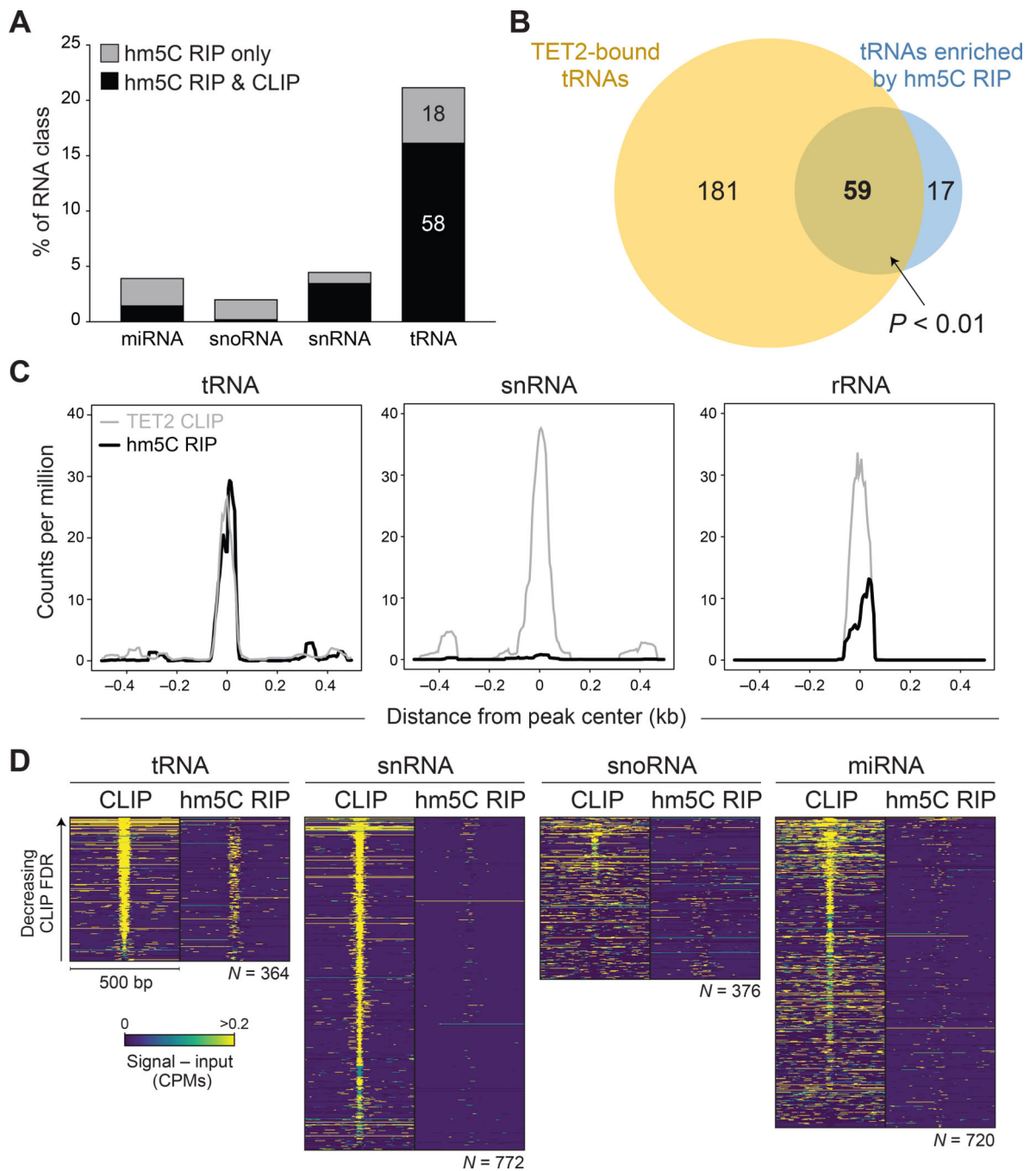
Figure 3. Distribution of hm5C and TET2 binding overlap within tRNAs

(A) The bars indicate the proportion of small RNAs from each of the four indicated classes enriched by hm5C RIP as determined by the presence of a peak with FDR $< 10^{-5}$. The subset of each class containing an overlapping TET2 CLIP peak is shown in black, whereas the RNAs only containing a hm5C RIP peak are shown in gray.

(B) Overlap of tRNAs containing a TET2 CLIP peak (left) with those containing a hm5C RIP peak (right). The *P*-value is from the hypergeometric distribution.

(D) Heatmaps for CLIP and hm5C RIP signals for the indicated small RNA classes after removing the respective input. The heatmaps were sorted according to the FDR of CLIP enrichment.

Data are from two independent immunoprecipitations.

A

B

C

Transfected construct

IB: HA

D

E

Coomassie

F
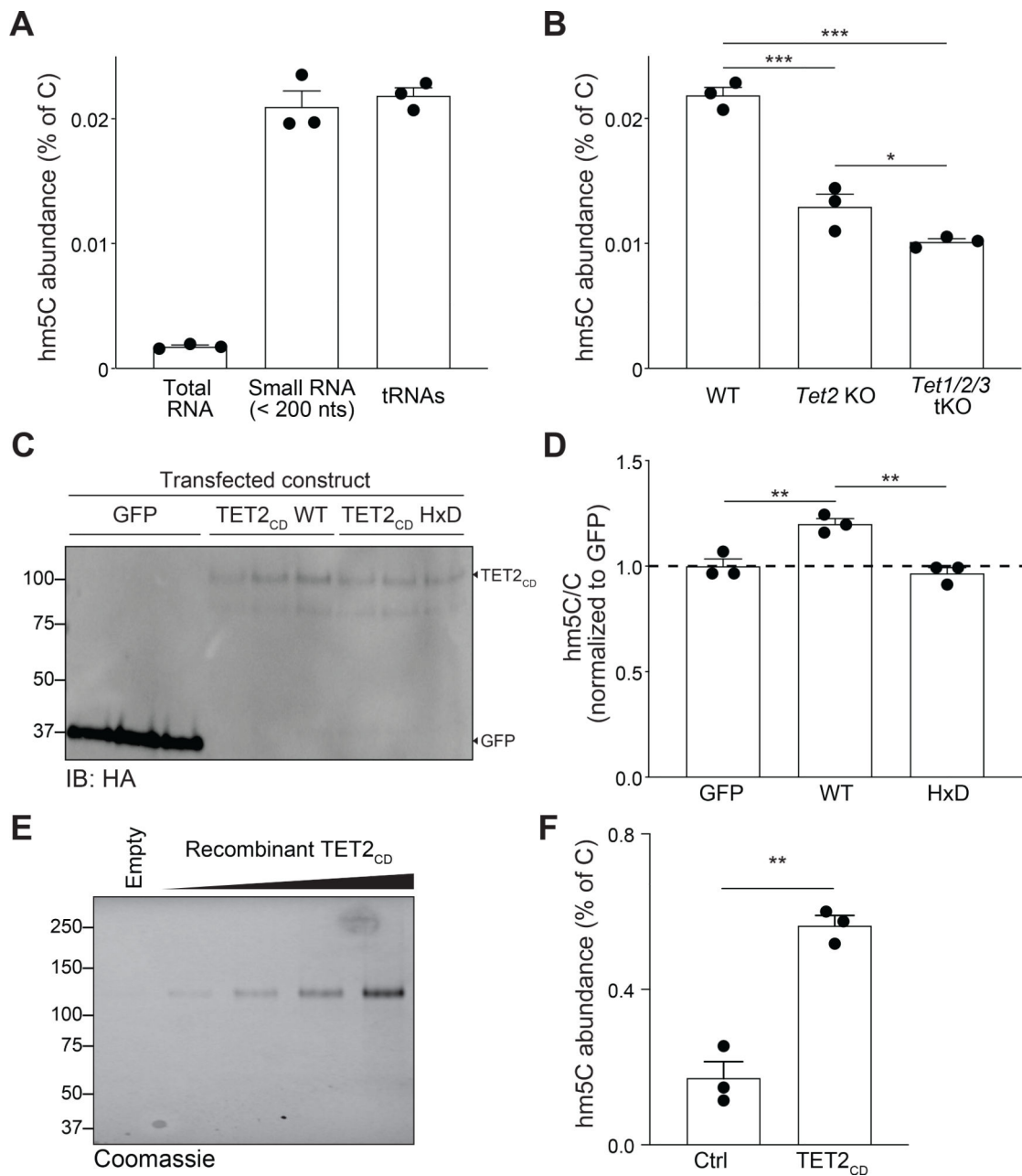
Figure 4. hm5C is enriched on tRNAs and depleted in the absence of TETs

(A) Mass spectrometry quantification of hm5C in total RNA (left), size-selected RNA smaller than ~200 nts (middle), and gel-purified tRNAs (right) extracted from mESCs. Abundance of hm5C is expressed as % of unmodified cytidine. Replicates are from three independent cell cultures and RNA extractions.

(B) Mass spectrometry quantification of hm5C in gel-purified tRNAs from WT (left) *Tet2* single KO (middle) and *Tet1/2/3* triple KO (right) mESCs. Replicates are from three independent cell cultures and RNA extractions.

(C) Western blot showing expression levels of transiently transfected GFP, a C-terminal fragment of TET2 comprising the catalytic domain (TET2$_{CD}$ WT), and the same fragment

carrying inactivating mutations (TET2$_{CD}$ HxD). All transfected proteins were fused with the HA epitope tag.

(D) Mass spectrometry quantification of hm5C on tRNAs from *Tet1/2/3* triple KO cells transfected with GFP, TET2$_{CD}$ WT, or TET2$_{CD}$ HxD. The bars represent the hm5C/C ratio normalized to the GFP control. Replicates are from three independent transfections for each construct.

(E) Coomassie staining of the purified mouse TET2 fragment used in (F).

(F) Mass spectrometry quantification of hm5C on tRNAs purified from *Tet1/2/3* triple KO cells and incubated with purified recombinant mouse TET2 catalytic domain fragment. All components except the enzyme were included in the control reaction. Replicates are from three independent reactions.

Bars represent mean + s.e.m. ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$. *P*-values are from one-way ANOVA followed by Holm-Sidak test.

Uncropped blot images for C and E are shown in Supplementary Figure 1.
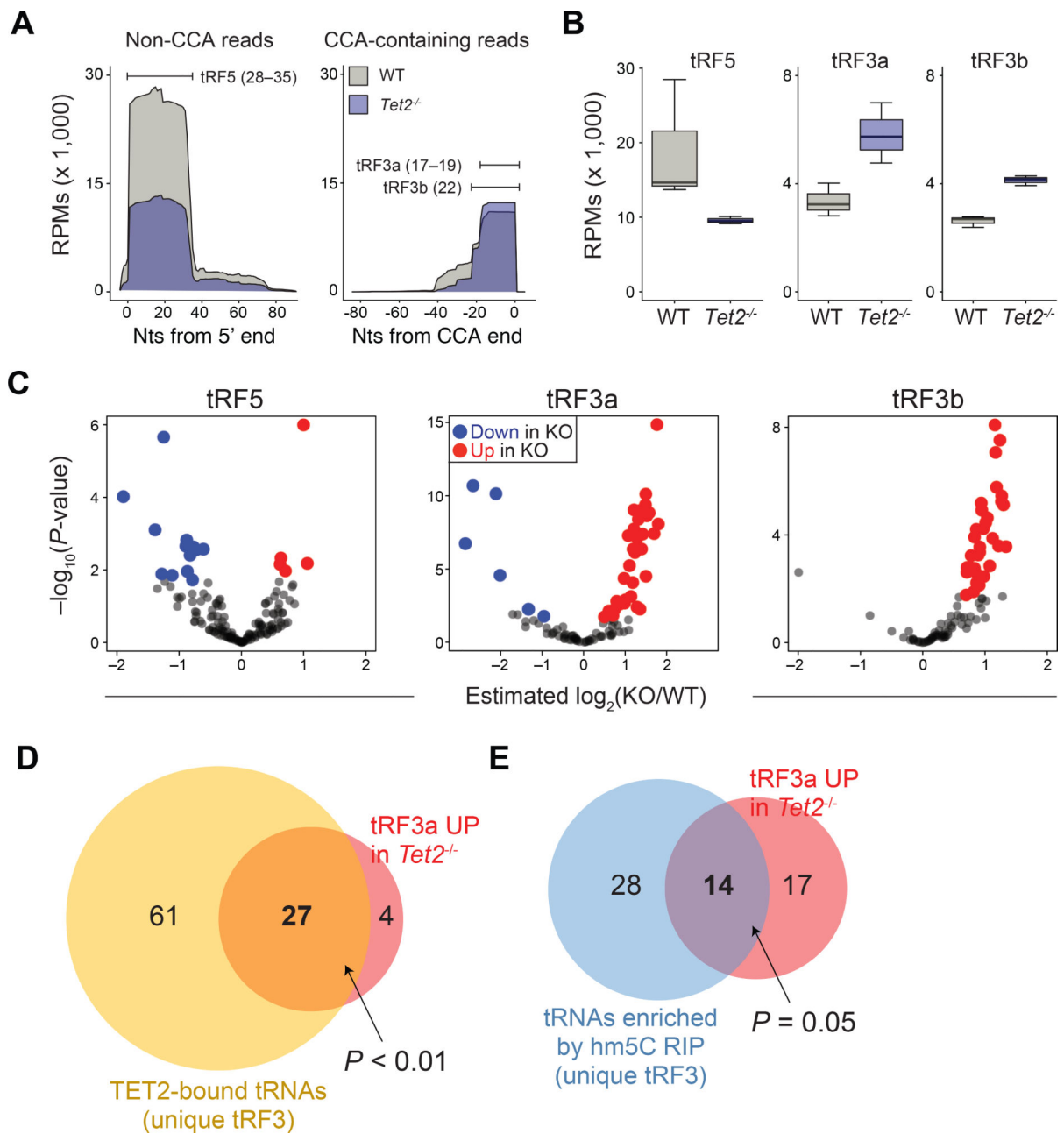
Figure 5. Loss of TETs affects the balance between classes of tRFs

(A) Coverage of tRNA genes by non-CCA (left) and CCA-containing (right) reads in small RNAs purified from WT or *Tet2⁻/⁻* cells. Plots show the average reads per million mapped (RPMs). Position of the three types of tRFs discussed in the text are indicated (tRF5, tRF3a, and tRF3b).

(B) Quantification of (A) but only considering size-filtered reads; 28–35 for tRF5, 17–19 (inclusive of CCA) for tRF3a, and 22 (inclusive of CCA) for tRF3b. The box blot elements are defined as center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range.

(C) Differential expression analysis for individual tRFs in E14 and *Tet2*⁻/⁻ cells. Estimated (DESeq2) fold changes are plotted on the *x* axis and the log-converted *P*-value on the *y* axis. Blue and red dots highlight individual tRFs that pass an adjusted *P*-value cutoff of 0.1 and are downregulated or upregulated in the KO, respectively.

(D) Overlap of TET2-bound tRNAs as determined by CLIP (Fig. 2), and the tRF3a significantly upregulated in *Tet2*⁻/⁻ cells as determined in (C). TET2-bound tRNAs were grouped according to the predicted sequence of the tRF3 produced from them. The *P*-value was calculated based on the hypergeometric distribution. Only the 31 upregulated tRF3a that were also detected in either CLIP or input samples were considered.

(E) Same as (D) but showing the overlap with tRNAs enriched by hm5C RIP (Fig. 3). Replicates are from three independent cell cultures and RNA purifications per genotype.