nature portfolio

Peer Review File

The extensive m5C epitranscriptome of Thermococcus kodakarensis is generated by a suite of RNA methyltransferases that support thermophily

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
After the first excitement as to how epitranscriptomes could affect a wide range of cellular processes, the fledgling field of epitranscriptomics has encountered various technical roadblocks with implications as to the validity of early epitranscriptomics mapping data. For instance, the low specificity of (supposedly) modification-specific antibodies for the enrichment of modified RNAs, has been ignored for too long and is only now recognized for its dismal reproducibility (between different labs). Furthermore, early attempts to map individual epitranscriptomes using sequencing-based techniques are largely characterized by the deliberate avoidance of orthogonal approaches aimed at confirming the existence of RNA modifications that have been originally identified by sequencing.
Improved methodology, the inclusion of various controls, and better mapping algorithms as well as the application of robust statistics for the identification of false-positive RNA modification calls have allowed revisiting original (seminal) publications whose early mapping data allowed making hyperbolic claims about the number, localization and importance of RNA modifications, especially in mRNA. Besides the existence of m6A in mRNA, the detectable incidence of RNA modifications in mRNAs has drastically dropped.
As for m5C, the subject of the manuscript submitted by Fluke et al., its identification in mRNA goes back to Squires et al., 2012 reporting on >10.000 sites in mRNA of a human cancer cell line, followed by intermittent findings reporting on pretty much every number between zero to > 100.000 m5C sites in different human cell-derived mRNA transcriptomes. The reasons for such discrepancies are likely of a technical nature. Importantly, all studies reporting on actual transcript numbers that were modified relied on RNA bisulfite sequencing, an NGS-based method, that can discriminate between methylated and non-methylated Cs after chemical deamination of C but not m5C. The method has a notoriously high background due to deamination artifacts, which occur largely due to incomplete denaturation of double-stranded regions (denaturing-resistant) of RNA molecules. Furthermore, m5C sites in mRNAs have now been mapped to regions that have not only sequence identity but also structural features of tRNAs. Various studies revealed that the highly conserved m5C RNA methyltransferases NSUN2 and NSUN6 do not only accept tRNAs but also other RNAs (including mRNAs) as methylation substrates, which in combination account for most of the m5C sites in human mRNA transcriptomes.
Given the generally low abundance and sub-stoichiometry of many internal mRNA modifications, it stands to reason how a few transcripts containing a particular RNA modification would exert biological impact among a majority of unmodified transcripts with the same sequence identity. To answer such questions, epitranscriptomics needs to become more quantitative, and divert from the unscientific trust in the notion that everything that we can detect is also biologically meaningful.
In light of this, the manuscript by Fluke et al., is reporting on “The extensive and dynamic m5C epitranscriptome of Thermococcus kodakarensis is generated by a suite of RNA methyltransferases that support life in the extremes.”
This work mapped m5C at nucleotide resolution in a model hyper-thermophile archaeal organism grown under laboratory conditions using RNA bisulfite sequencing. The authors annotated potential m5C writer enzymes, performed in vitro methylation assays, and used genetic manipulation to remove single-copy genes alone or in combination to pinpoint the substrate specificity of these enzymes. Finally, archaeon strains with impaired m5C writer activities displayed limited growth under hyper-thermophilic conditions, indicating that m5C in RNA is required for life under such conditions.
The manuscript is well suited to contribute important information to the RNA modification community. Unfortunately, the authors chose to exclude to query both tRNA and rRNAs for m5C modification. Importantly, in light of the more recent findings that m5C in eukaryotic mRNA is catalyzed mostly by NSUN2 and NSUN6, two tRNA methyltransferases, and occurs largely at tRNA-like sequences (at much fewer mRNAs than previously thought; PMID: 31061524, 33330931, 34691665), an obvious question is if the few mRNA positions reported to be bisulfite conversion-refractory in the mRNA of this archaeon are representing technical artifacts, or are the consequence of Star activity of a particular m5C RNA modification circuitry, which evolved to modify rRNAs and tRNAs but also takes aim at similar sequence or structure in other RNAs. To approach an answer to the latter question, this

reviewer asks the authors to implement some additional bioinformatics analysis, before publication of the manuscript in Nature Communications can be recommended. The additional analysis would help the field to generalize if prokaryotes, just like eukaryotes, have a propensity to allow mRNA modifications in sequence or structural context akin to non-coding RNAs,
General comments:
1) It remains unclear why the authors dismiss the modification of rRNAs and tRNAs with m5C and focus their work on mRNA. Importantly in the author's reading, the use of the word "epitranscriptome" seems to only apply to mRNA modifications. Hence, the m5C epitranscriptome is only that which is related to mRNAs. However, the few sites in mRNA that are reproducibly (3/3 experiments) bisulfite-refractory are only a minor fraction of the expressed mRNAs. Calling those few potential modification sites, not only the epitranscriptome but also extensive, seems overly confident. If the authors would like to correct that view, they should change the title of the manuscript to accommodate the notion, that they only analyzed mRNAs, and that the m5C status of mRNAs is potentially very low in this organism.
2) The authors should consider changing their wording with respect to modified cytosines. As it stands, RNA bisulfite sequencing reveals potential modification sites as bisulfite-refractory. Since this study represents a kind of de novo assembly of an unknown m5C epitranscriptome, not every cytosine after bisulfite treatment must have originated from a methylated cytosine. Detected cytosines might be technical artifacts or could have been the results of other RNA modifications, which interfere with bisulfite-mediated deamination of cytosine. In this respect, have the authors considered that a cytosine modification such as N4-acetylation (ac4C), which has reported roles in RNA stabilization, could be revealed by RNA bisulfite sequencing? In any case, it would be prudent to not call every bisulfite-refractory cytosine a methylated cytosine, unless proven with orthogonal technology.
3) tRNAs and rRNAs are the most abundantly modified RNA species in any cell type. Modifications affect the folding, maturation, stability, and function of both abundant non-coding RNAs. Given that the respective RNA modification enzymes have evolved for the purpose of ensuring tRNA and rRNA functionality, it is likely that these enzymes are testing and re-testing various RNA substrates for modification, including mRNAs. Hence, it is likely that these enzymes, like other (processive) enzymes, do (aberrantly) modify mRNAs, which are not their perfect substrates. It is therefore conceivable that the low stoichiometry that has been generally observed for m5C in mRNAs in eukaryotes (a given mRNA identity contains about 20% modified Cs at one particular position), is the result of such a scanning plus stochastic star activity of the respective modification enzymes.
In light of the low levels reported as bisulfite-refractory at particular positions in specific mRNAs, it would be interesting to address whether these sites represent tRNA-like or rRNA-like structures. Can the authors compare the positions that are bisulfite-refractory in the few mRNAs with the sequences and structures of archaeal rRNAs and tRNAs (TKt41 and 16S-tRNAAla-23S)?
4) This reviewer does not see a reason why the authors would include the results showing bisulfite-refractory cytosines in 2 out of 3 repeats at all. Preferably, many more repeats should have been performed to arrive at robust data, which, as nicely shown by McIntyre et al., Scientific Reports (2020), increases the reproducibility of m6A detection in mRNA. If the data for 3/3 repeats exist, what is the use of excluding one repeat unless the authors try to prop up the numbers of potentially modified mRNAs from 77 to 232?
5) Have the authors repeated RNA bisulfite sequencing of a particular locus that was predicted to be methylated by their high-throughput sequencing experiments? The perceived message of the manuscript is that only 77 mRNAs contained reproducible detectable cytosines after bisulfite treatment (3/3 experiments). However, the stoichiometry at a particular nucleotide seems to differ, averaging at 15-20%. Unless the authors designed the study including UMIs in the sequencing runs (the mentioning of which is nowhere to be found), it would be important to test for some of the positions in the 77 mRNAs, and ask if a locus is really bisulfite-refractory on more than a few mRNA molecules that were successfully reverse-transcribed into cDNA. This could strengthen or weaken the message by showing that the identified loci are indeed containing non-converted cytosines outside of deamination-resistant nucleotide sequences.
6) Sodium bisulfite treatment of RNA causes major degradation. Creating cDNA libraries from such RNAs with reduced base complexity (presence of AGU with very low levels of Cs remaining) will cause

biases during cDNA synthesis. For instance, poly-U stretches due to deamination of cytosine within CUCU context might affect the efficiency of reverse transcriptase creating poly-A. In addition, poly-A stretches might cause PCR bias when amplifying cDNAs. The authors should explain, how they determined that reads covering a specific genomic (transcript) region have been derived from many cDNA molecules and therefore from multiple bisulfite-treated RNAs. Or, they should state how they excluded mapped reads derived from a limited number of cDNA molecules that represent only a few "surviving" RNAs. This is especially important when looking at the non-converted positions in reads with high coverage. Since the authors do not mention barcoding of cDNA synthesis to address the potential for cDNA synthesis or PCR bias, performing this post-hoc exercise should allow the reader to gauge the extent of bias in the data sets.
7) Related to the previous comment, the authors should provide information about the expression level of those mRNAs, which they flag as containing m5C. If those mRNAs are highly expressed, because they encode proliferation-relevant proteins and contain sequence features that mimic tRNAs or rRNAs, they might become substrates of some m5C circuitry and therefore appear to be methylated. In addition, any such mRNA identity that is represented by many individual molecules will "survive" the deamination reaction better than lowly expressed mRNA identity. This might uncover particular mRNAs, which appear to be more methylated just because they are highly expressed in the archaeon. The authors should try to address this possibility of skewing the read-out, which would then indicate or refute the existence of an artifact introduced by the applied methodology.
8) The authors use CGmaptools to calculate C/T coverage. Does the software map cytosines in other contexts than CpG? And can they elaborate on why they did not use software that was designed to map RNA bisulfite sequencing data?
9) In light of the small genome of Thermococcus kodakaransis, have the authors considered addressing the extent of bisulfite artifacts by transcribing the genome and using the "naked" RNA as a reference for their bisulfite sequencing? This approach has been described in pmid: 34594034.
10) Related to the sub-stoichiometry point, and given the low methylation levels for most of the identified m5C sites in mRNA in eukaryotes (PMID: 31061524, 33330931, 34691665), it would be prudent to alert readers that even in the event of losing a particular m5C site in the respective RNA modification enzyme knockout strain, the jury is still out there to prove beyond doubt that m5C at a particular mRNA position (in 1-2 out of 10 mRNAs, or for the "high coverage" mRNA, 5 out of 100) is biologically meaningful, especially under extreme conditions such as high temperatures.
Specific comments:
Abstract: "While many modifications are limited in frequency, restricted to non-coding RNAs, or present only in select organisms, 5-methylcytidine (m5C) is abundant across diverse RNAs and fitness-relevant across Domains of life…"
> This reviewer asks to revisit the generalized statement. Currently, m5C is restricted to non-coding RNAs, and the few mRNA positions reported appear to be shrinking with improved technological repeat experiments. If one gives those data the benefit of the doubt, m5C in mRNA is about one magnitude lower than m6A, which is considered to be the most abundant with 1-3 modified Adenosines per average mRNA.
Page 4: "Modifications to even individual residues in long RNAs are known to elicit dramatic impacts on structure14–16, stability17–20, protein-interactions21, cellular…"
> Reference 15 and 16 refer to work on tRNAs, which are not long RNAs. Reference 21 is a review. Please, cite primary data.
Page 5: "Specialized nucleotides within rRNAs and tRNAs often provide essential stability and importantly, these modified RNAs must be passed to daughter cells, providing a heritable role for epitranscriptomic modifications."
> It is unclear, why modified RNAs must be passed to daughter cells. Please, provide primary references.
Page 5: "Evidence suggests this regulatory paradigm extends to mRNAs and across Domains, with RNA modifications being essential for all life"
> Most RNA modifications are not essential for life as shown in numerous mutagenesis screens in yeast and other organisms.
Page 5: "Site-specific, and often sub-stoichiometric, modification of mRNA is linked to core biological

functions and responses to environmental changes."
> This statement cannot be a general one. Exactly the sub-stoichiometric modification is one of the great conundrums in RNA modification research. As of now, nobody has ever tested how many site-specific modifications are required for biological impact. Removing the writer enzymes in their entirety is insufficient to answer this question. Hence, stoichiometry of RNA modifications has not been linked to core biological functions.
Page 6: "We identified at least 232 unique m5C sites in a diverse set of coding and non-coding RNAs, established that ~10% of all unique transcripts contain m5C, and that mRNA represents the largest fraction of m5C-modified RNAs…"
> This statement should read that the authors identified at least 77 unique sites refractory to bisulfite in 3 out of 3 experiments. Furthermore, that mRNA represents the largest fraction of m5C-modified RNAs is only supported because rRNA and tRNA were depleted from the input RNA into the bisulfite sequencing. Hence, this statement is misleading and will therefore likely be picked up by AI-generated abstracts produced by paper mills, which will reiterate that the m5C is only found in mRNA in this organism.
Page 13: "Using samtools, alignments were removed from bam files where MAPQ score was < 20 and when detected as a PCR duplicate…"
> How did the authors identify PCR duplicates with that method?
Page 16: "Visual inspection indicates that modification frequencies level-off at ~10%, indicating many modifications with a frequency below 10% may be false-positives. We determined that a high-confidence m5C site must reach a 10% modification frequency."
> What is visual inspection of high-throughput data?
> Why is a high confidence site determined by a 10% and not by a 90% modification frequency?
Page 21-22: "All 23 nt RNA substrates were ordered from IDT and resuspended in nuclease-free water"
> How did the authors arrive at this sequence length?
Page 25: "…hydroxymethylcytidine, 3-methyl cytidine, or 4-methyl cytidine, by comparing retentions times and mass transitions of corresponding modified nucleosides (data not shown)."
> Given that this organism is likely harboring more modifications than just m5C, it would be commendable to show the detected nucleosides?
> The depiction of the mass spec data in the main Figure is insufficient. Normally mass spectra of chromatographic peaks with retention times are shown with m/z on the x-axis and the relative abundance of the modification on the y-axis (as provided in the supplementary data).
Page 25: "RNA sequencing libraries were prepared for total RNA and mRNA-enriched fractions ..."
> It is unclear how the authors enriched for mRNA. All they did according to M & M is to deplete rRNAs or remove small RNAs (<200 nt) from total RNA.
Page 26: "Owing to the depth of high-quality sequencing, such as that of sites with ≥ 1000x coverage, the acceptable minimum m5C frequency was lowered to 5%."
> It is unclear why higher coverage should allow for including positions that are bisulfite-refractory in 5/100 transcripts.
Page 29: "We mapped m5C within GCG codons (alanine, n=18) to an extent that is fourfold higher than expected if by random chance. Modifications to codons GCU (alanine, n=13), GGC (glycine, n=18), CCG (proline, n=24), and CUG (leucine, n=16) were similarly enriched. When codons include multiple cytidines, there is a strong bias for which cytidine is selected for modification"
> High CG content has previously been connected to deamination artifacts. Could the authors provide secondary structure predictions of these sites within these mRNAs (similar to Figure 4H)? If these codons form secondary structures with neighboring sequences, then the creation of bisulfite artifacts is likely.
Page 30: "Taken together, these data indicate a strong positional bias in m5C sites in particular codon and amino acid contexts, likely indicating m5C impacts the translatability of these codons."
> m5C does not affect base pairing but more stacking interactions. How would one m5C-modified cytosine in 5 out of 100 mRNAs affect the translation of the encoded protein resulting in biological impact, especially under hyper-thermophile conditions?
Page 32: "To our surprise, certain strains deleted for putative m5C-specific and alternative RMTase

enzymes showed gains in m5C sites or increased modification frequencies."
> What is an increased modification frequency? Please, provide these data to the reader. Unless these observations can be repeated with targeted RNA bisulfite sequencing and barcoded cDNA synthesis to control for RNA molecule input and, hence number of cDNAs, this result could be interpreted as a sign of deamination artifacts, which are amplified by RNaseH (-) RT and PCR.
Page 37: "Structural comparisons between the three RNA substrates (Figure 4H) adumbrate that the secondary and tertiary structures between these RNAs may play a role in the substrate recognition or catalytic activity of rTK2304 in vitro…"
> What is "adumbrate"?
Page 38: "To confirm modification of the TK1911 mRNA at the site predicted based on the loss of in vivo modification due to deletion of TK2304, the 101 nt substrate containing a C or U at the central position was mixed with unlabeled SAM and rTK2304 in vitro, purified, then digested to single nucleosides for LC-MS/MS analysis. We observed 0.37% of m5C/C ratio (or 5% oligo modification frequency) on the C-containing 101-nt RNA (Figure 4F and G)"
> Given that the enzyme TK2304, when acting in situ, might not methylate all substrates for various reasons, it remains unclear why a clean enzyme presented with a clean RNA is not able to methylate more than 5% of the position. If one takes any classic rRNA or tRNA methylating enzyme, those will modify pretty much every transcript at the target nucleotide position. The result of the in vitro methylation activity indicates that the methylation frequency is similar to the methylation background that was defined by the authors for the analysis of the bisulfite sequencing data. And, therefore, the results agree with the notion that this enzyme is likely not an enzyme acting on mRNA, hence cannot be called a robust enzyme (as stated by the authors in a downstream paragraph).

Reviewer #2:
Remarks to the Author:
In their manuscript "The extensive and dynamic m5C epitranscriptome of Thermococcus kodakarensis is generated by a suite of RNA methyltransferases that support life in the extremes" Fluke at al. investigate the complexity of the m5C epitranscriptome in T. kodakarensis in the context of requirements for survival in hyper thermophilic environments.

The manuscript was extremely pleasant to read. I notably appreciate how the authors rigorously evaluate artefacts, in particular in bisulfite sequencing and discuss negative findings. In addition, a lot of features are validated orthogonally through in vitro methylation assays and mass spectrometry. This approach renders these datasets highly informative and not limited to research on archaea but applicable to all domains of life. The manuscript is quite complete, since it covers the mapping at single base m5c in T. kodakarensis epitranscriptome up to the identification of the responsible enzymes, the validation of the modification sites through the use of mutants for the methyltransferases and finally to the function of these modifications in extreme environments. The manuscript is well written, and I have only minor concerns to be addressed. I support publication of the study in Nature Communications.

Minor points.

In table 2 it would be informative to indicate which type of RNA (tRNA, rRNA, mRNA …etc.) is lost or gained at m5c sites. This will help to understand the specificity of each enzyme. A table or a better visualization of the sites lost and gained in double mutants would be also very useful.

A .txt file or. xlxs table giving the coordinates of the 232 high-confidence and reproducible m5C sites (page 32 row 736) and adding detailed information regarding the type of RNA they are found in, the position, from which R5CMT these have been added etc., would be useful for future users of the datasets.

Reviewer #3:
Remarks to the Author:
Comments to the authors:

This manuscript (457436_0_art_file_765223 and 765225) reports the results of epitranscriptome analysis of m5C in Thermococcus kodakarensis. The authors have newly identified five RNA methyltransferase genes for m5C modifications in coding and non-coding RNAs. Furthermore, the authors have found that some m5C methyltransferases are required for efficient growth of T. kodakarensis. To date, seven modifications (m5s2U54, m1A58, m7G46, archaeosine15, m22G10, ac4C at multiple positions and phosphorylation of U47) in tRNA have been reported to be required for efficient growth of thermophiles at high temperatures. As far as this reviewer knows, this is the first report of the requirement of m5C modification in RNA at high temperatures for efficient growth of thermophiles. In addition, previously reported modifications are tRNA modifications. If rRNA and/or mRNA m5C modification(s) are essential for efficient growth of T. kodakarensis at high temperatures, the findings by the authors are quite novel. Therefore, this reviewer believes that this study significantly contributes to RNA modification field and that this manuscript includes many scientific merits to be published. However, this reviewer would like to ask the authors revision of the manuscript and addition of experiments.

Major Points:
(1) Please reconsider the descriptions in INTRODUCTION. In the first section, the authors mainly claim the importance of RNA modifications in eukaryotes (mesophiles). In the second section, the authors describe the significance of mRNA modifications (mesophiles). However, the current research by the authors mainly focuses on the m5C modification in rRNA and mRNA from thermophilic archaeon. General readers in Nature Communications probably understand the importance of RNA modification. Therefore, the authors should explain the significance of epitranscriptome analysis of thermophiles more adequately. As describe above, seven modifications in tRNA have been reported to be required for efficient growth of thermophiles at high temperatures. Please see these references in additions to references 16 and 2 (Page 5 lines 102-103).
Droogmans L. et al. (2003) Nucleic Acids Res. 31, 2148-2156.
Shigi N. et al. (2006) J. Biol. Chem. 281, 2104-2113.
Tomikawa C. et al. (2010) Nucleic Acids Res. 38, 942-957.
Hirata A. et al. (2019) J. Bacteriol. 201, e00448-19.
Ohira T. et al. (2022) Nature 605, 372-379.
In the case of T. kodakarensis, m5s2U54, m22G10 and m1A58 stabilize the tRNA structure and are required for efficient growth at high temperatures in addition to archaeosine15 and ac4C at multiple positions.

(2) This reviewer understands that the authors mainly focus on long RNAs. However, this report is very important for tRNA modification field as well as rRNA and mRNA modification fields. Furthermore, if identified m5C methyltransferases do not act on tRNA, it shows the significance of rRNA and mRNA m5C modifications at high temperatures. Therefore, this reviewer strongly recommends the authors to add some experiments to clarify whether identified m5C methyltransferases act on tRNA or not. The authors have gene deletion strains and recombinant enzymes. Therefore, the authors can check the methyl-transfer activity of each enzyme towards tRNA fraction from the corresponding gene deletion strain. (This reviewer does not request purification of tRNA molecular species and identification of methylation site(s) by MS. This reviewer understands the difficulty of these experiments.)
Furthermore, please add information concerning m5C modifications in tRNA. For example, the authors detected only two m5C modifications in tRNA (Figure 1C and Page 28 lines 626-628). Are these positions 48 and 49 in tRNA? Moreover, Page 35 line 793. “Tkt41 (tRNATrp) persisted through sequencing, and TK0360 was identified to methylate this transcript.” Transfer RNATrp possesses three m5C modifications (m5C32, m5C48 and m5C49: see Hirata A. et al. (2019) J. Bacteriol. 201, e00448-

19). The responsible enzyme for m5C32 has not been identified yet. If Tk0360 gene product acts on C32 in tRNATrp, Tk0360 gene product is a novel tRNA methyltransferase. Please check the modification site.

(3) As described in the Discussion, m5C modification does not disturb the formation of Watson-Crick base pair. Therefore, the methyl group in m5C probably contributes to strengthen the hydrophobic interaction (stacking effect) in stem structures. Although the authors discuss the codon positions of m5C modification through the manuscript, the stacking effect may be more important for decoding process on ribosome at high temperatures. (This is only my comment. If the authors agree with this idea, please add descriptions).

(4) Page 43 lines 985. "Previous studies have also shown that most m5C sites where the R5CMT has been identified are present in a sequence context of no more than ~4 nucleoside."
In the reference 50, the m5C modification site in mRNA from S. solfataricus has been identified as AUC*GANGU (* shows the methylation site). Please see Page 7 and Figure 5 in reference 50. The discussion should be altered.

(5) Data interpretation of methylation of Tk1911 mRNA. Clarification of substrate RNA recognition mechanism of m5C RNA methyltransferases is very difficult. This phenomenon was firstly reported by yeast tRNA m5C methyltransferase (Trm4). See Motorin and Grosjean (1999) RNA 5, 1105-1118. Similar phenomenon has been reported. In the experiments by the authors, molecular ratio of enzyme and substrate is 1:1. Under this condition, Tk1911 23 nt may be methylated efficiently. If necessary, please add reference and discussion. Furthermore, if possible, add the annotation of Tk1911 gene product.

(6) This report is very important for rRNA modification field. Add detailed positions in rRNAs (modification sites of m5C in rRNAs and identified modification sites of each m5C methyltransferase in this study).

Minor points
(1) Fonts are not unified in several parts in main text. For example, page 15 lines 334-340.
(2) Figures do not appear according to the orders because several figures are explained in Materials and Methods section. For example, Page 15 line 340, Figure 2B.
These descriptions are friendly to general readers. However, those may not match the journal style.
(3) Table 1. Function unknown RNA methyltransferase-like genes and already-known RNA methyltransferase genes exist in T. kodakarensis genome except for genes in this list. Does this list show putative RNA methyltransferases used in this study?

Reviewers:

The three reviews of our original manuscript were positive, supportive of its publication in *Nature Communications*, and the suggestions for modifications to such resulted in an improved, clarified, and more impactful revised manuscript.

Language has been added that clarify the function and fitness-relevance of m^5C modifications to coding sequences may be a result of star activity of methyltransferase circuitry. Supplementary figure 1 has been expanded to include m/z mass spectrometry data at the request of reviewer 1. Additional experiments that measure the activity of methyltransferases on small RNA fractions have been performed at the request of reviewer 3. The introduction has been expanded and now includes a summary of what is known about hyperthermophilic epitranscriptomes. Additional bioinformatics analysis regarding RNA structure was performed at the request of reviewer 1 and 2. All grammatical and typographical errors were corrected.

In addition to supplying the revised manuscript, we provide both a marked version of the original manuscript (to highlight the positions and extent of changes) and a point-by-point response to each comment from the reviewers. All authors have contributed to the changes that improve the manuscript, and all authors approve the revised manuscript. No changes to authorship were warranted.

-Tom Santangelo (on behalf of all authors)

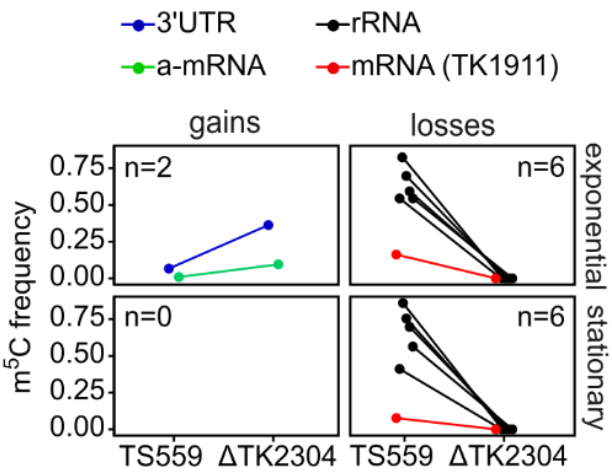
Reviewer #2

In their manuscript "The extensive and dynamic m^5C epitranscriptome of Thermococcus kodakarensis is generated by a suite of RNA methyltransferases that support life in the extremes" Fluke at al. investigate the complexity of the m^5C epitranscriptome in T. kodakarensis in the context of requirements for survival in hyper thermophilic environments. **The manuscript was extremely pleasant to read.** I notably appreciate how the authors rigorously evaluate artifacts, in particular in bisulfite sequencing and discuss negative findings. In addition, a lot of features are validated orthogonally through in vitro methylation assays and mass spectrometry. This approach renders these datasets highly informative and not limited to research on archaea but applicable to all domains of life. **The manuscript is quite complete**, since it covers the mapping at single base m^5C in T. kodakarensis epitranscriptome up to the identification of the responsible enzymes, the validation of the modification sites through the use of mutants for the methyltransferases and finally to the function of these modifications in extreme environments. The manuscript is well written, and I have only minor concerns to be addressed. **I support publication of the study in Nature Communications.**

We thank the reviewer for their feedback on the manuscript and support for its publication in *Nature Communications*. Our exhaustive efforts to call high confidence m^5C sites and eliminate false positives had not gone unnoticed by this reviewer. We strive to be as accurate, clear, and concise in our verbiage as possible, and were pleased to know that this reviewer found our manuscript well written. We have made several improvements to the manuscript in response to each point of concern.

(1) In table 2 it would be informative to indicate which type of RNA (tRNA, rRNA, mRNA …etc.) is lost or gained at m^5C sites. This will help to understand the specificity of each enzyme. A table or a better visualization of the sites lost and gained in double mutants would be also very useful.

The type of RNA and modification frequency for each m^5C site that is lost/gained as a result of the deletion of an R5CMT is illustrated in figure 4C (depicted below) and supplementary figure 9A-D.iii. The color corresponds to the RNA type (UTR, rRNA, mRNA, etc) and the modification frequency in parent strain TS559 and deletion strain (ΔTK2304) of each site is graphed on the y-axis. The data is faceted by whether the change in modification frequency was a gain or loss. Two growth conditions were evaluated (exponential and stationary growth). We realize this detail was originally not mentioned in the text, and we have modified the manuscript to point this out to the reader: Sites where m^5C was gained or lost were mapped mainly to rRNAs (Figure 4C, **black lines**), but one m^5C site was detected in a single mRNA (Figure 4C, **red line**)



(C) The change in modification frequency between parent strain TS559 and ΔTK2304 is faceted by growth phase and direction of regulation. The number of sites (n) is listed within each window and the color corresponds to RNA type.

(2) A .txt file or. xlxs table giving the coordinates of the 232 high-confidence and reproducible m5C sites (page 32 row 736) and adding detailed information regarding the type of RNA they are found in, the position, from which R5CMT these have been added etc., would be useful for future users of the datasets.

Supplementary file 1 has been provided and details genomic coordinates of each m^5C site detected and sites gained/lost in each deletion strain. Each site is exhaustively annotated with information including the modification frequency, total coverage, and m^5C coverage in each replicate as well as the type, gene name, length, etc. of the RNA in which the m^5C site is found. We realize that this was not clearly stated in the original text. *The main text has been modified to make this explicitly clear to the reader.* The text now reads: “The genomic coordinates and complete annotation of each m^5C site is recorded in Supplementary File 1.”

Reviewer #3

This manuscript (457436_0_art_file_765223 and 765225) reports the results of epitranscriptome analysis of m5C in Thermococcus kodakarensis. The authors have newly identified five RNA methyltransferase genes for m5C modifications in coding and non-coding RNAs. Furthermore, the authors have found that some m5C methyltransferases are required for efficient growth of T. kodakarensis. To date, seven modifications (m5s2U54, m1A58, m7G46, archaeosine15, m22G10, ac4C at multiple positions and phosphorylation of U47) in tRNA have been reported to be required for efficient growth of thermophiles at high temperatures. As far as this reviewer knows, this is the first report of the requirement of m5C modification in RNA at high temperatures for efficient growth of thermophiles. In addition, previously reported modifications are tRNA modifications. If rRNA and/or mRNA m5C modification(s) are essential for efficient growth of T. kodakarensis at high temperatures, the findings by the authors are quite novel. Therefore, **this reviewer believes that this study significantly contributes to RNA modification field and that this manuscript includes many scientific merits to be published.** However, this reviewer would like to ask the authors revision of the manuscript and addition of experiments.

We thank the reviewer for their feedback on the manuscript and support for its publication in *Nature Communications*. This is indeed the first reported evidence that m^5C supports growth under hyperthermophilic conditions. We sincerely appreciate the reviewer's recognition of the novelty and merit of this report. We have made several improvements in response to each point of concern.

(1) Please reconsider the descriptions in the introduction. In the first section, the authors mainly claim the importance of RNA modifications in eukaryotes (mesophiles). In the second section, the authors describe the significance of mRNA modifications (mesophiles). However, the current research by the authors mainly focuses on the m5C modification in rRNA and mRNA from thermophilic archaeon. General readers in Nature Communications probably understand the importance of RNA modification. Therefore, the authors should explain the significance of epitranscriptome analysis of thermophiles more adequately. As described above, seven modifications in tRNA have been reported to be required for efficient growth of thermophiles at high temperatures. Please see these references in additions to references 16 and 2 (Page 5 lines 102-103).

Droogmans L. et al. (2003) Nucleic Acids Res. 31, 2148-2156.
Shigi N. et al. (2006) J. Biol. Chem. 281, 2104-2113.
Tomikawa C. et al. (2010) Nucleic Acids Res. 38, 942-957.
Hirata A. et al. (2019) J. Bacteriol. 201, e00448-19.

Ohira T. et al. (2022) Nature 605, 372-379.
In the case of T. kodakarensis, m5s2U54, m22G10 and m1A58 stabilize the tRNA structure and are required for efficient growth at high temperatures in addition to archaeosine15 and ac4C at multiple positions.

This is an excellent suggestion. These references have been added and the following paragraph has been incorporated into the introduction:

"Many archaea thrive in conditions inhospitable to most extant life. Maintaining RNA structure at high temperature, or at the extremes of salinity, pressure, or pH are facilitated, in part, by chemical modifications. Specific RNA modifications are critical for life at high temperatures, and loss of epitranscriptomic modifications is often lethal.[37–40] In *Thermus thermophilus*, a hyperthermophilic bacteria, tRNA 1-methyladenosie (m^1A)[39], 7-methylguanosine (m^7G)[41], *and* 2-Thioribothymidine[38] are required for growth at high temperatures. Likewise, *in Thermococcus kodakarensis, a hyperthermophilic archae*on, 2-dimethylguanosine (m^2_2G)[40] and 2'O-phosphouridine (p^2U) in tRNAs are required for growth at high temperatures. Not only are individual modified residues in tRNAs essential for hyperthermophilic growth, 4-acetylcytidine (ac^4C) and 2'O-methyl-ac4C in *Pyrococcus furiosus* and ac4C in *T. kodakarensis* are markedly increased with rising growth temperature.[2,42] 5-methyl-2-thiouridine (m^5s^2U), m^2_2G, archaeosine (G+) and m^1A stabilize *T. kodakarensis* tRNA structure and promote hyperthermophilic growth.[16,37,40]"

(2) This reviewer understands that the authors mainly focus on long RNAs. However, this report is very important for tRNA modification field as well as rRNA and mRNA modification fields. Furthermore, if identified m5C methyltransferases do not act on tRNA, it shows the significance of rRNA and mRNA m5C modifications at high temperatures. Therefore, this reviewer strongly recommends the authors to add some experiments to clarify whether identified m5C methyltransferases act on tRNA or not. The authors have gene deletion strains and recombinant enzymes. Therefore, the authors can check the methyl-transfer activity of each enzyme towards tRNA fraction from the corresponding gene deletion strain. (This reviewer does not request purification of tRNA molecular species and identification of methylation site(s) by MS. This reviewer understands the difficulty of these experiments.) …

We thank the reviewer for this comment and agree that m^5C in tRNA is biologically significant. Supplementary figure 1C (small fraction) shows a robust signal for m^5C in RNAs < 200nt, which includes tRNA. Although not fully supported in this work, we strongly hypothesize that m^5C is present in tRNAs and the R5CMTs identified here likely do act on tRNAs. Unfortunately, tRNAs are very difficult to sequence and are not captured well in our sequenced libraries despite repeated attempts. Due to methodological limitations, we can not include a comprehensive analysis of m^5C in

tRNAs at this time, although a comprehensive analysis of tRNA m^5C is underway. Current efforts are geared towards establishing OTTR-seq for the sequencing and modification analysis of tRNAs in *T. kodakarensis*.

We agree that simple assays can potentially establish whether the suite of RNA methyltransferases may target tRNAs. Bulk assays of total small RNA preparations, likely dominated by tRNAs, with radiolabel SAM transfer driven by *in vitro* activities of purified enzymes were therefore carried out with five recombinant enzymes.

In support of likely tRNA targeting activity, rTK2122 shows considerable activity (defined as transfer of radiolabeled methyl groups from SAM to the bulk RNA population) on small (tRNA) and large (rRNA) RNA fractions derived from the strain ΔTK2122 but not the parent strain, indicating that the protein product of gene TK2122 likely targets tRNAs and rRNAs.

rTK2304 shows significant activity on the large RNA fraction (but not small fractions) derived from ΔTK2304 but not from the parent strain. These data would suggest TK2304 encodes a methyltransferase that targets rRNA, which is supported by the BS-seq data.

We were unable to detect rTK1935 activity on either small or large RNA pools, which is interesting since we are able to detect high activity on small synthetic oligos. We speculate that the pool of RNA isolated for these experiments is dominated by mature r- and tRNA, and the TK1935 protein product may act co-transcriptionally, relying more heavily on nucleotide sequence rather than structure. This may explain the more complex sequence motif detected at m^5C sites lost in ΔTK1935.

We were not able to detect rTK0360 or rTK0872 activity on RNA fractions or on synthetic RNAs, likely indicating these enzymes require additional factors that we have not identified (such as RNA secondary/tertiary structures that are not easily achieved *in vitro*).

We have opted to not include this data in the manuscript pending thorough investigation. It is unclear whether rTK2304, rTK1935, rTK0360, or rTK0872 do not show activity on the small RNA fraction due to our *in vitro* reactions conditions, whether other *in vivo* factors may be necessary for activity, or if these enzymes truly do not target tRNAs.

… Furthermore, please add information concerning m^5C modifications in tRNA. For example, the authors detected only two m^5C modifications in tRNA (Figure 1C and

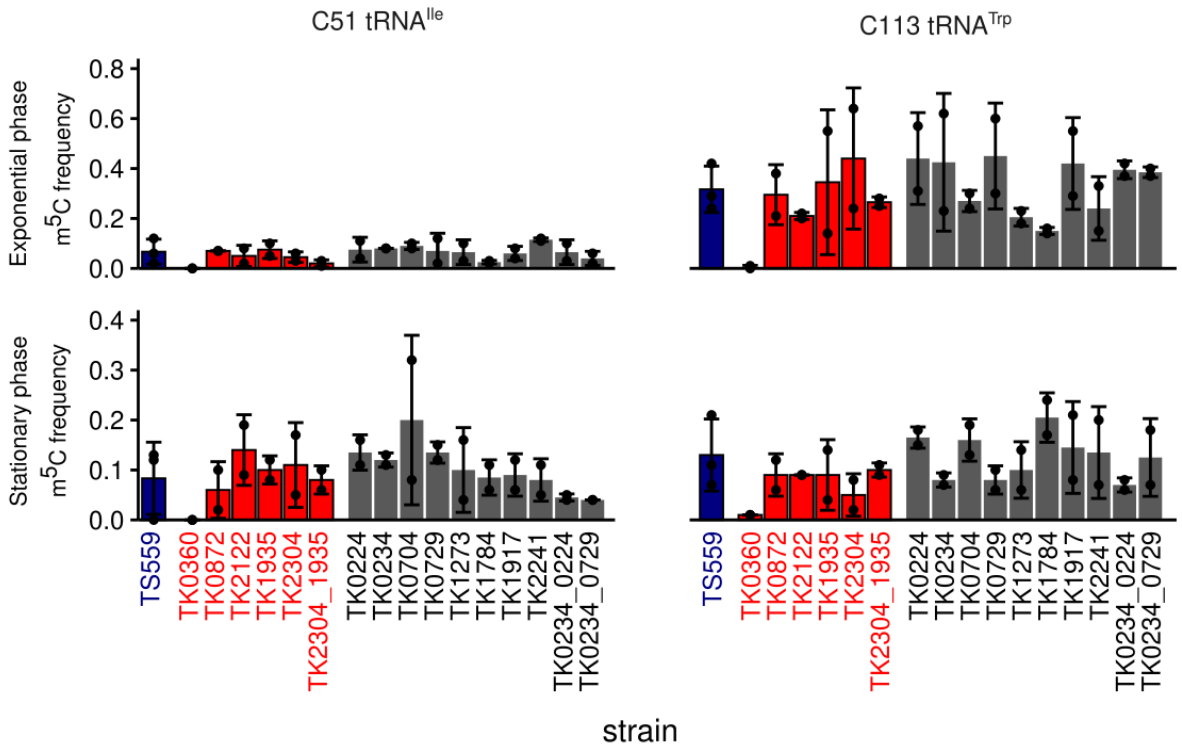
Page 28 lines 626-628). Are these positions 48 and 49 in tRNA? Moreover, Page 35 line 793. “Tkt41 (tRNATrp) persisted through sequencing, and TK0360 was identified to methylate this transcript.” Transfer RNATrp possesses three m5C modifications (m5C32, m5C48 and m5C49: see Hirata A. et al. (2019) J. Bacteriol. 201, e00448-19). The responsible enzyme for m5C32 has not been identified yet. If Tk0360 gene product acts on C32 in tRNATrp, Tk0360 gene product is a novel tRNA methyltransferase. Please check the modification site.

We indeed detected two m^5C sites in two tRNAs, TKt30 ($tRNA^{Ile}$) and TKt41 ($tRNA^{Trp}$), in our TS559 strain and all of our deletion strains, apart from ΔTK0360 **(See figure below)**. As our library prep protocol excludes RNAs < 200 nt, we are confident that these tRNAs were captured in their immature form BEFORE the m^5C modification profile is fully realized. **The text has been modified to clarify this point.**

$tRNA^{Ile}$ is transcribed in an operon with another TKt31, resulting in an initial transcript that is slightly greater than 200 nt. During the exponential growth phase, position C51 is modified at or below our high confidence thresholds in all strains but ΔTK0360. During the stationary phase, modification frequencies at C51 increase substantially, and this site is occupied at or above our thresholds for nearly all strains except ΔTK0360. It is likely that the modification frequency of C51 in the mature TKt31 tRNA is more stable across replicates and strains and that the product of gene TK0360 is responsible for modifying this tRNA. However, given the stochasticity of the data at this m^5C site, we can not be confident that TK0360 is modifying C51.

TKt^{Trp} (TKt41) is 140 nt in its mature form and likely includes longer 3’ and/or 5’ ends that are further processed post-transcriptionally. The initial tRNA transcript is likely around 200 nt or greater. This tRNA is modified at C113, and the modification frequency is more stable across replicates and strains in this tRNA, regardless of growth phase. In ΔTK0360, this m^5C site is definitively lost, and we can say with reasonable confidence that the product of gene TK0360 is responsible for modifying this tRNA at C113.

Further experimentation is needed to establish the complete and mature tRNA modification profiles and the enzymes responsible for such.



(3) As described in the Discussion, m^5C modification does not disturb the formation of Watson-Crick base pair. Therefore, the methyl group in m^5C probably contributes to strengthening the hydrophobic interaction (stacking effect) in stem structures. Although the authors discuss the codon positions of m5C modification through the manuscript, the stacking effect may be more important for decoding process on ribosomes at high temperatures. (This is only my comment. If the authors agree with this idea, please add descriptions).

We certainly agree with this idea, and we have added it to the discussion section.

(4) Page 43 lines 985. “Previous studies have also shown that most m5C sites where the R5CMT has been identified are present in a sequence context of no more than ~4 nucleoside.” In the reference 50, the m5C modification site in mRNA from *S. solfataricus* has been identified as AUC*GANGU (* shows the methylation site). Please see Page 7 and Figure 5 in reference 50. The discussion should be altered.

While it is true that in *S. solfataricus*, m^5C sites were present in a more complex sequence motifs, we stand by our original statement that most m^5C sites previously reported are found in ~4 nt sequence contexts. In this manuscript, m^5C sites lost in TK1935 deletion strain appear in complex sequence motifs as well. However, most m^5C sites correlated with the loss of one of the other enzymes (TK2304, TK2122, and

TK0872) are found in a 3-4 nt sequence contexts. The text has been modified to clarify this point.

(5) Data interpretation of methylation of Tk1911 mRNA. Clarification of substrate RNA recognition mechanism of m^5C RNA methyltransferases is very difficult. This phenomenon was firstly reported by yeast tRNA m^5C methyltransferase (Trm4). See Motorin and Grosjean (1999) RNA 5, 1105-1118. Similar phenomenon has been reported. In the experiments by the authors, the molecular ratio of enzyme and substrate is 1:1. Under this condition, Tk1911 23 nt may be methylated efficiently. If necessary, please add reference and discussion. Furthermore, if possible, add the annotation of Tk1911 gene product.
The manuscript has been modified to clarify these points. We've also added the statement, "Gene TK1911 encodes a hypothetical protein with an unknown function."

(6) This report is very important for rRNA modification field. Add detailed positions in rRNAs (modification sites of m^5C in rRNAs and identified modification sites of each m5C methyltransferase in this study).
We fully agree with this comment. Complete analysis of rRNA modification frequencies in each strain and the genomic coordinates are recorded in supplementary file 3. *The main text has been modified to make this explicitly clear to the reader.* The text now reads: "Genomic coordinates and modification frequencies in the 16S-$tRNA^{Ala}$-23S rRNA operon are recorded in Supplementary File 3."

Minor points
(1) Fonts are not unified in several parts in main text. For example, page 15 lines 334-340.
The text font has been changed such that all fonts are Arial size 11.

(2) Figures do not appear according to the orders because several figures are explained in Materials and Methods section. For example, Page 15 line 340, Figure 2B.
These descriptions are friendly to general readers. However, those may not match the journal style.
The Materials & Methods section has been moved to the end, per the formatting requirements.

(3) Table 1. Function unknown RNA methyltransferase-like genes and already-known RNA methyltransferase genes exist in T. kodakarensis genome except for genes in this list. Does this list show putative RNA methyltransferases used in this study?
Table 1 and Supplementary Table 1 initially include a list of all putative RNA methyltransferases that we hypothesized to encode for R5CMTs. The "Description" in

these tables are bioinformatic predictions according to Uniprot and KEGG, but their functions have not been experimentally probed. RNA methyltransferases that we show in this study to install m^5C are moved to Table 1.

Reviewer #1

After the first excitement as to how epitranscriptomes could affect a wide range of cellular processes, the fledgling field of epitranscriptomics has encountered various technical roadblocks with implications as to the validity of early epitranscriptomics mapping data. For instance, the low specificity of (supposedly) modification-specific antibodies for the enrichment of modified RNAs, has been ignored for too long and is only now recognized for its dismal reproducibility (between different labs). Furthermore, early attempts to map individual epitranscriptomes using sequencing-based techniques are largely characterized by the deliberate avoidance of orthogonal approaches aimed at confirming the existence of RNA modifications that have been originally identified by sequencing.

Improved methodology, the inclusion of various controls, and better mapping algorithms as well as the application of robust statistics for the identification of false-positive RNA modification calls have allowed revisiting original (seminal) publications whose early mapping data allowed making hyperbolic claims about the number, localization and importance of RNA modifications, especially in mRNA. Besides the existence of m6A in mRNA, the detectable incidence of RNA modifications in mRNAs has drastically dropped.

As for m5C, the subject of the manuscript submitted by Fluke et al., its identification in mRNA goes back to Squires et al., 2012 reporting on >10.000 sites in mRNA of a human cancer cell line, followed by intermittent findings reporting on pretty much every number between zero to > 100.000 m5C sites in different human cell-derived mRNA transcriptomes. The reasons for such discrepancies are likely of a technical nature. Importantly, all studies reporting on actual transcript numbers that were modified relied on RNA bisulfite sequencing, an NGS-based method, that can discriminate between methylated and non-methylated Cs after chemical deamination of C but not m5C. The method has a notoriously high background due to deamination artifacts, which occur largely due to incomplete denaturation of double-stranded regions (denaturing-resistant) of RNA molecules. Furthermore, m5C sites in mRNAs have now been mapped to regions that have not only sequence identity but also structural features of tRNAs. Various studies revealed that the highly conserved m5C RNA methyltransferases NSUN2 and NSUN6 do not only accept tRNAs but also other RNAs (including mRNAs) as methylation substrates, which in combination account for most of the m5C sites in human mRNA transcriptomes.

Given the generally low abundance and sub-stoichiometry of many internal mRNA modifications, it stands to reason how a few transcripts containing a particular RNA

modification would exert biological impact among a majority of unmodified transcripts with the same sequence identity. To answer such questions, epitranscriptomics needs to become more quantitative, and divert from the unscientific trust in the notion that everything that we can detect is also biologically meaningful.

In light of this, the manuscript by Fluke et al., is reporting on “The extensive and dynamic m5C epitranscriptome of Thermococcus kodakarensis is generated by a suite of RNA methyltransferases that support life in the extremes.” This work mapped m5C at nucleotide resolution in a model hyper-thermophile archaeal organism grown under laboratory conditions using RNA bisulfite sequencing. The authors annotated potential m5C writer enzymes, performed in vitro methylation assays, and used genetic manipulation to remove single-copy genes alone or in combination to pinpoint the substrate specificity of these enzymes. Finally, archaeon strains with impaired m5C writer activities displayed limited growth under hyper-thermophilic conditions, indicating that m5C in RNA is required for life under such conditions.

The manuscript is well suited to contribute important information to the RNA modification community. Unfortunately, the authors chose to exclude to query both tRNA and rRNAs for m5C modification. Importantly, in light of the more recent findings that m5C in eukaryotic mRNA is catalyzed mostly by NSUN2 and NSUN6, two tRNA methyltransferases, and occurs largely at tRNA-like sequences (at much fewer mRNAs than previously thought; PMID: 31061524, 33330931, 34691665), an obvious question is if the few mRNA positions reported to be bisulfite conversion-refractory in the mRNA of this archaeon are representing technical artifacts, or are the consequence of Star activity of a particular m5C RNA modification circuitry, which evolved to modify rRNAs and tRNAs but also takes aim at similar sequence or structure in other RNAs. To approach an answer to the latter question, this reviewer asks the authors to implement some additional bioinformatics analysis, before publication of the manuscript in Nature Communications can be recommended. The additional analysis would help the field to generalize if prokaryotes, just like eukaryotes, have a propensity to allow mRNA modifications in sequence or structural context akin to non-coding RNAs,

We thank the reviewer for their feedback on the manuscript and general support for its publication in *Nature Communications*. We recognize and largely agree that RNA modification circuitry in many cases has evolved to target tRNA and rRNAs, and although not supported in this work, we can not discount the possibility that “off target” modifications provided an evolutionary advantage and may in fact drive fitness. Further research is necessary to determine if m^5C in coding sequences is biologically

meaningful and drives fitness or if these modifications are simply “mistakes”. This study serves to provide a foundation to address such questions. We have added a qualifying statement to the introduction and discussion “...How m^5C residues impact the fate and functions of mRNAs at the individual transcript level has been historically challenging to address due to low abundance RNAs and substoichiometric modification frequencies in conventional model organisms. Given the generally low abundance and sub-stoichiometry of many internal mRNA modifications, it stands to reason how a few transcripts containing a particular RNA modification would exert biological impact among a majority of unmodified transcripts. It remains contested whether mRNAs are specifically targeted for modification and whether m^5C in coding regions are functionally relevant…”

General comments:

1) It remains unclear why the authors dismiss the modification of rRNAs and tRNAs with m5C and focus their work on mRNA. Importantly in the author’s reading, the use of the word “epitranscriptome” seems to only apply to mRNA modifications. Hence, the m5C epitranscriptome is only that which is related to mRNAs. However, the few sites in mRNA that are reproducibly (3/3 experiments) bisulfite-refractory are only a minor fraction of the expressed mRNAs. Calling those few potential modification sites, not only the epitranscriptome but also extensive, seems overly confident. If the authors would like to correct that view, they should change the title of the manuscript to accommodate the notion, that they only analyzed mRNAs, and that the m5C status of mRNAs is potentially very low in this organism.

We apologize for the misunderstanding that we only analyzed mRNA. We analyzed RNA >200 nt by BS-seq and a variety of RNA size fractions by mass spectrometry (Supplementary Figure 1). We did not exclude rRNA. In fact, we included an extensive analysis of candidate m^5C in rRNA. Figure 3 and Supplementary File 3 are entirely dedicated to m^5C occurrence in the 16S-tRNA-23S rRNA operon, and we did not detect m^5C in either copy of the 5S rRNA (nor are there any cryoEM densities at cytidines that would be indicative of methylation at C5). We mapped 21 m5C sites in mature ribosomal RNA, correlated the loss of m^5C sites with the R5CMT responsible for its installation, and conclude that the number of m^5C sites is ~10X more abundant in *T. kodakarensis* rRNA compared to humans and *E. coli* rRNA.

We performed RNA Bisulfite sequencing on RNAs > 200 nt. Although not rigorously supported in this work, we believe that m^5C is present in tRNAs and the R5CMTs identified here likely do act on tRNAs. Supplementary Figure 1C shows that we are detecting a robust m^5C signal by mass spectrometry in the small RNA fraction, which includes tRNAs. Unfortunately, **tRNAs are very difficult to sequence and are not captured well in our sequenced libraries after repeated attempts.** Due to

methodological limitations, we can not at this time include a comprehensive analysis of m^5C in tRNAs, but current efforts are geared towards establishing OTTR-seq for the sequencing of tRNAs. We state in the results and discussion sections that the m^5C epitranscriptome presented in this manuscript surely underestimates the true number of m^5C sites in the transcriptome and that additional sites are likely present in tRNAs.

Our use of the word “extensive” refers to the relative extent of m^5C incorporation into the epitranscriptome. In Supplementary Figure 1, we show that the [m^5C]/[C] ratio is ~25X higher compared to that in human cell lines. In Figure 1A, we show that m5C is incorporated into about 10% of unique and expressed RNAs via BS-seq. We will highlight here that *T. kodakarensis* encodes only 2,306 genes and more than half are expressed under our laboratory growth conditions. So, 77 or 232 m5C sites constitutes a relatively large fraction of unique and expressed RNAs. We also show that m^5C is present in diverse RNAs, including mRNA. Based on the high relative levels of m^5C incorporation into the transcriptome and the wide-range of RNA species in which m^5C is found (regardless of whether we consider 2/3 or 3/3 replicates), we believe that “extensive” is the appropriate word to describe the m^5C epitranscriptome in *T. kodakarensis*, and we respectfully disagree that m^5C occurrence is low.

2) The authors should consider changing their wording with respect to modified cytosines. As it stands, RNA bisulfite sequencing reveals potential modification sites as bisulfite-refractory. Since this study represents a kind of de novo assembly of an unknown m^5C epitranscriptome, not every cytosine after bisulfite treatment must have originated from a methylated cytosine. Detected cytosines might be technical artifacts or could have been the results of other RNA modifications, which interfere with bisulfite-mediated deamination of cytosine. In this respect, have the authors considered that a cytosine modification such as N4-acetylation (ac^4C), which has reported roles in RNA stabilization, could be revealed by RNA bisulfite sequencing? In any case, it would be prudent to not call every bisulfite-refractory cytosine a methylated cytosine, unless proven with orthogonal technology.

We thank the reviewer for this comment. We strive to be as accurate as possible in the terminology we use. We have added a qualifying statement that reads, “Cytidines retained after bisulfite treatment are therefore presumed to be m^5C. However we can not rule out the possibility that some retained cytidines are instead occupied by other bisulfite resistant modifications or are otherwise resistant to deamination... We identified at least 232 candidate m^5C sites in a diverse set of coding and non-coding RNAs…”

hm^5C is also detectable to BS-seq, but we did not detect any traces of hm^5C above the limit of detection by mass spectrometry. We also failed to detect trances of m^4C or m^3C

in *Thermococcus* RNAs. We acknowledge and agree that other RNA modifications, even unknown modifications, may resist bisulfite-driven deamination. But, ac^4C (and m^4C) are not resistant to BS-deamination and BS-seq is not suitable for ac^4C detection. Recent work by us and colleagues has established methodologies for detecting ac^4C that rely on specific chemistries. Of the 404 ac^4C sites we mapped to the *T. kodakarensis* transcriptome **(PMID: 32555463)**, there is no overlap between the m^5C sites mapped here and the ac^4C sites mapped previously. This was a post-hoc analysis – we did not rely on the ac^4C data to inform our confidence measures of m^5C sites.

3) tRNAs and rRNAs are the most abundantly modified RNA species in any cell type. **Modifications affect the folding, maturation, stability, and function of both abundant non-coding RNAs.** Given that the respective RNA modification enzymes have evolved for the purpose of ensuring tRNA and rRNA functionality, it is likely that these enzymes are testing and re-testing various RNA substrates for modification, including mRNAs. Hence, it is likely that these enzymes, like other (processive) enzymes, do (aberrantly) modify mRNAs, which are not their perfect substrates. It is therefore conceivable that the low stoichiometry that has been generally observed for m^5C in mRNAs in eukaryotes (a given mRNA identity contains about 20% modified Cs at one particular position), is the result of such a scanning plus stochastic star activity of the respective modification enzymes. In light of the low levels reported as bisulfite-refractory at particular positions in specific mRNAs, it would be interesting to address whether these sites represent tRNA-like or rRNA-like structures. Can the authors compare the positions that are bisulfite-refractory in the few mRNAs with the sequences and structures of archaeal rRNAs and tRNAs (TKt41 and 16S-tRNAAla-23S)?

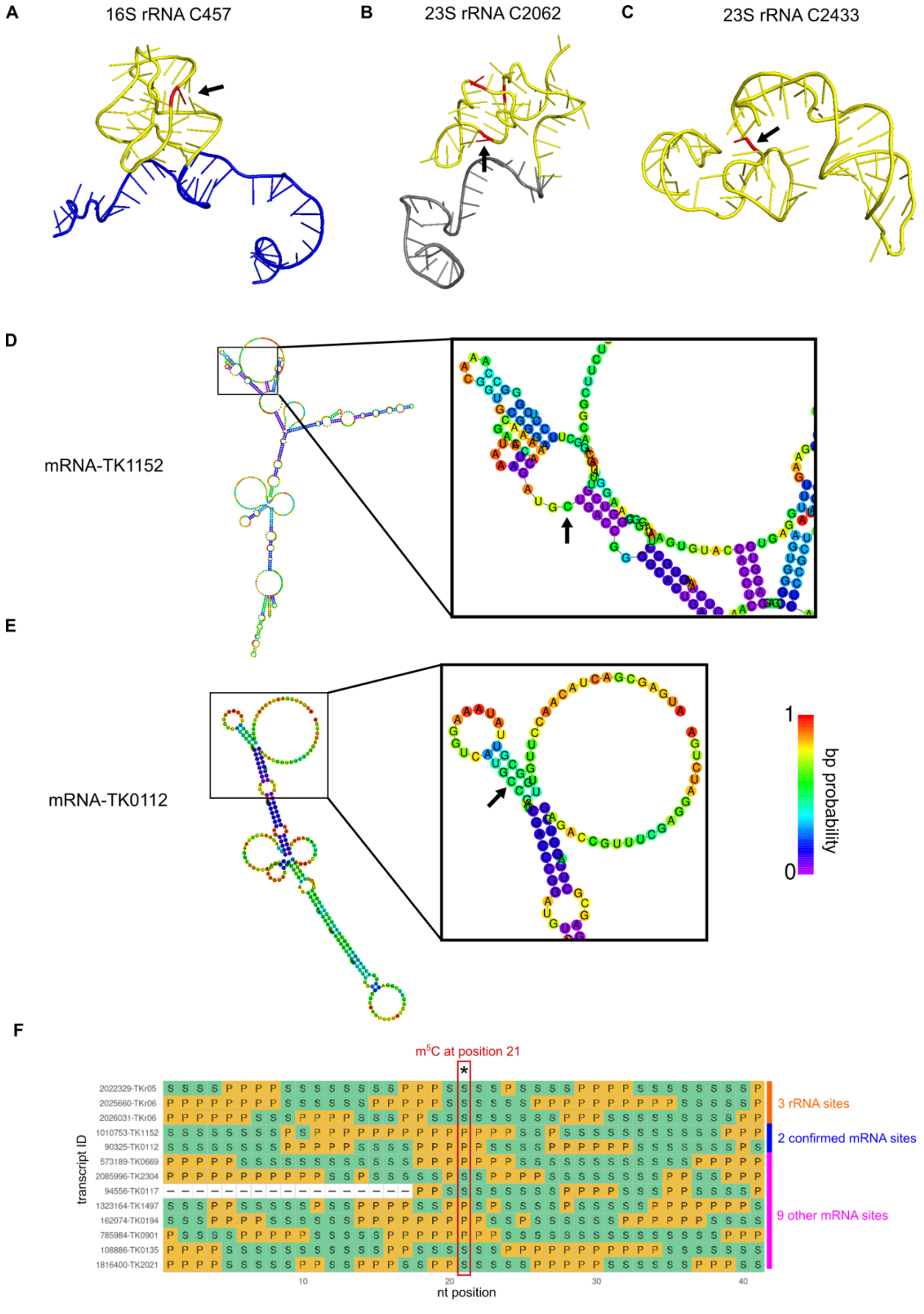
It is absolutely possible that mRNAs are “off-target” substrates for methylation where rRNA and tRNAs are the intended substrate. However we can not simply make this assumption; there have been very few studies that have probed whether m^5C mRNA modifications drive fitness. Decades of research has generated excellent evidence for the biological role of m^6A in alternative splicing of mRNAs. Such clear evidence that would suggest the biological role of m^5C in mRNA is lacking. However, it is clear that m^5C is abundant in Eukaryotic and Archaeal mRNAs. For what purpose is unknown. The manuscript presented here intends to provide the foundation for such experiments, and current efforts are geared towards understanding the fitness impacts of site-specific m^5C sites in coding sequences.

tRNA-like structures in mRNA being targeted for methylation was first reported for yeast Trm4 (PMID: 10445884). But, structural analysis of RNAs targeted for methylation is complicated. Below, we are providing 3 figures depicting our structural analysis of 3

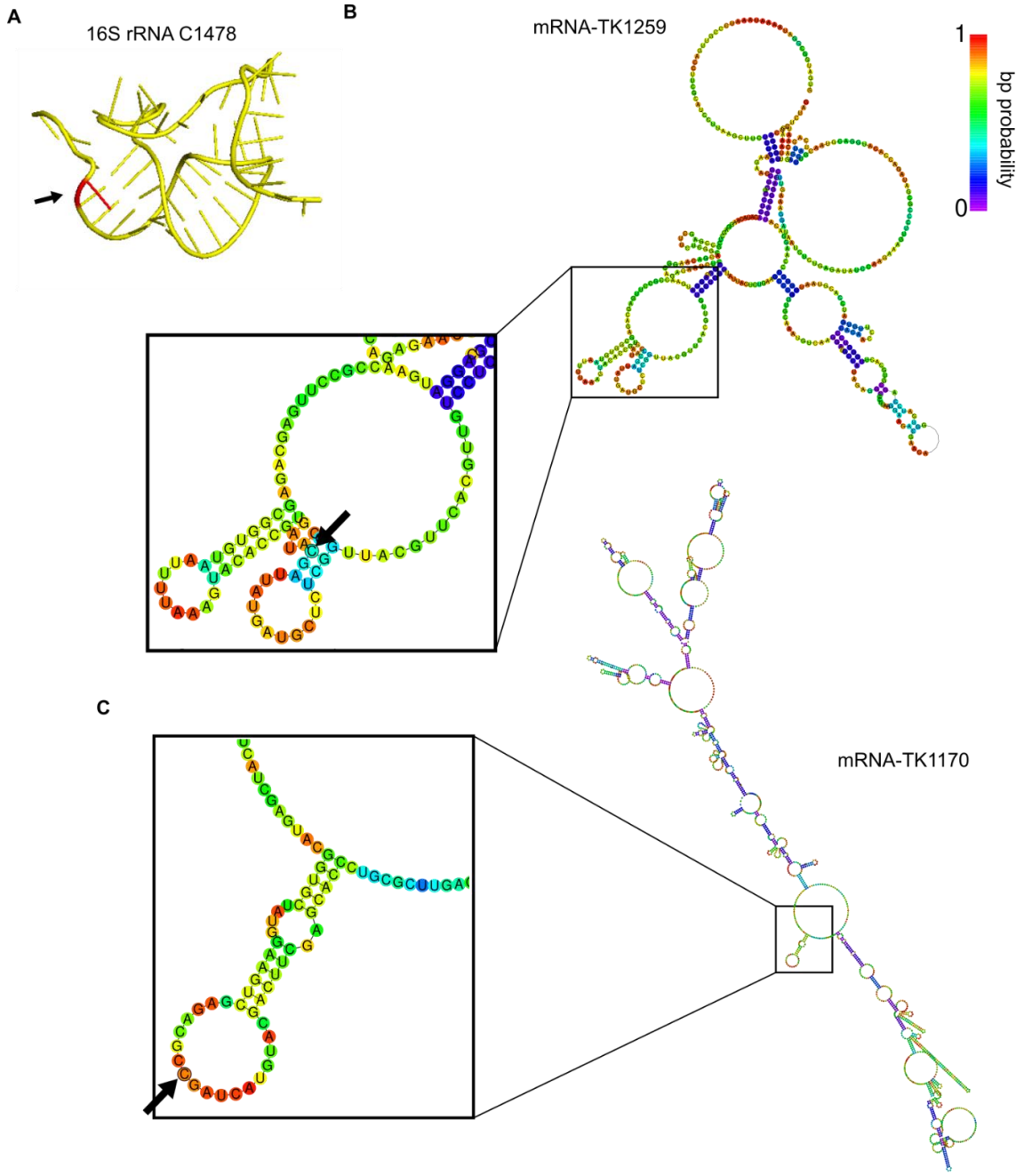
bona fide RNA m^5C methyltransferases reported here. We surveyed the available cryoEM structure of the *T. kodakarensis* ribosome for secondary and tertiary structures at m^5C sites as well as secondary structure predictions using vienna RNAfold. We anticipated to see structural similarities at least between sites that we confirmed by mass spectrometry or from the cryoEM structure. However, we did not see large consistencies in RNA structures targeted by each enzyme.

It is important to note that the predicted structures analyzed here are that of mature RNAs and do not represent intermediate/co-transcriptional structures that may be optimal targets for these enzymes. It is also extremely difficult to predict tertiary RNA structures (such as pseudoknots) which are known to play a role in targeting RNA binding enzymes. The predicted structures are also just that… predicted. It is unclear to us whether Vienna is accurately predicting structures at such high temperatures (85 degrees C), as the base pair probabilities (corresponding to nucleotide color) are not great. Given that emerging evidence suggests the *T. kodakarensis* epitranscriptome is densely modified with ac^4C (PMID: 32555463), structural predictions are also surely confounded by other modifications. The predicted secondary structures at m^5C sites indicate that roughly 50% of m^5Cs are base paired or single stranded, suggesting that the dataset is likely not contaminated with false positives due to denaturation resistance (please see later comments for more details).

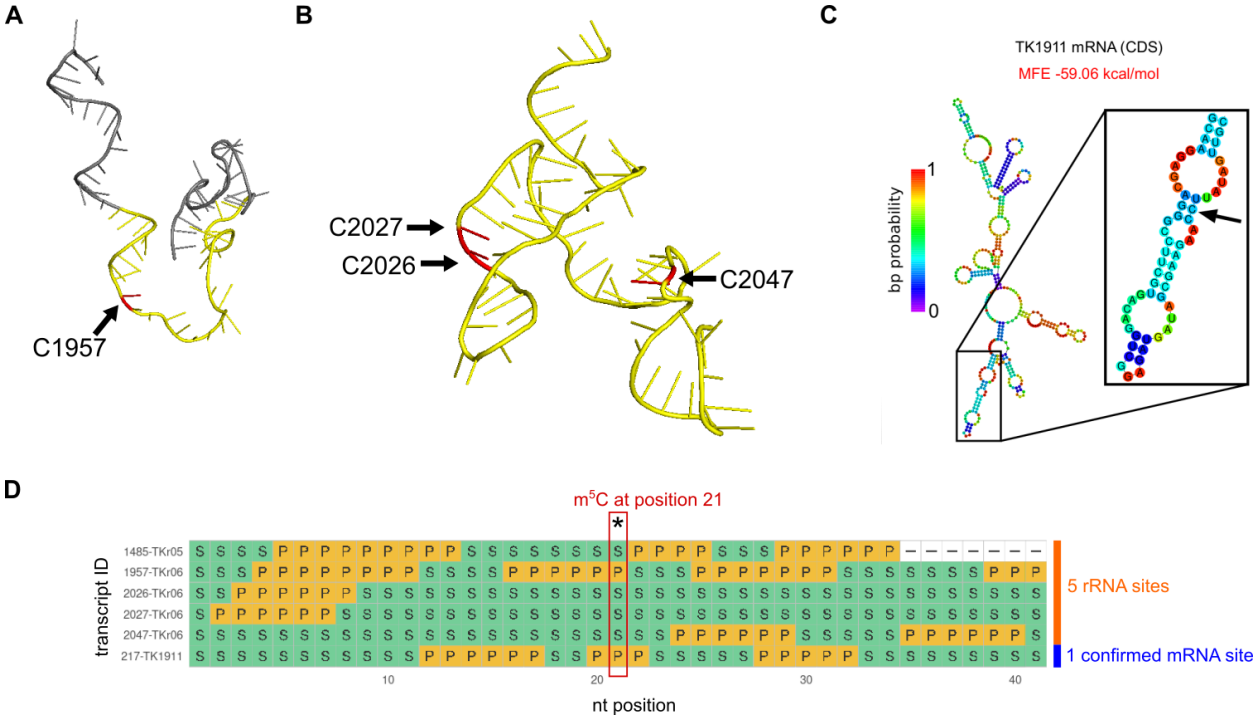
At this time, we are not confident that the structural predictions are truly reflective of structures at m^5C sites nor at the time of methylation. Perhaps a lab focused on such algorithms is better suited to identify structural targets.


A
16S rRNA C457
B
23S rRNA C2062
C
23S rRNA C2433
D
mRNA-TK1152
E
mRNA-TK0112
1
0
bp probability
F
m5C at position 21
transcript ID
2022329-TKr05
2025660-TKr06
2026031-TKr06
1010753-TK1152
90325-TK0112
573189-TK0669
2085996-TK2304
94556-TK0117
1323164-TK1497
162074-TK0194
785984-TK0901
108886-TK0135
1816400-TK2021
10
20
30
40
nt position
3 rRNA sites
2 confirmed mRNA sites
9 other mRNA sites

Structural analysis of RNAs targeted for methylation by the protein product of gene TK1935. (A-C) *The tertiary structure of rRNAs (CryoEM) where a loss in cytidine retention is detected (red, arrow) after RNA BS-seq of the strain deleted for gene TK1935.* ***(A)*** *C456 in the 16S rRNA and* ***(B)*** *C2062 and* ***(C)*** *C2433 in the 23S rRNA are double stranded.* ***(D, E)*** *The predicted structure of mRNAs that TK1935 methylate as demonstrated in vitro. The cytidine targeted for methylation is indicated by an arrow.* ***(F)*** *For each candidate m^5C site that was lost in the TK1935 deletion strain, the predicted secondary structure is aligned. "P" indicates paired and "S" indicates single stranded nucleotides. The cytidine targeted for methylation is at the center (position 21) and surrounded by 20 nt of up and downstream sequence. In the case of mRNA-TK0117, the m^5C site is the 4th nucleotide, and therefore only 3 nt upstream are included.*


A
16S rRNA C1478
B
mRNA-TK1259
1
0
bp probability
C
mRNA-TK1170
D
m5C at position 21
*
transcript ID
1478-TKr05
782-TK1170
177-TK1259
551-TK1022
362-TK0180
235-TK2201
819-TK0571
485-TK1111
483-TK2184
527-TK0308
317-TK2290
524-TK1946
2 rRNA sites
2 confirmed mRNA sites
9 other mRNA sites
10
20
30
40
nt position

Structural analysis of RNAs targeted for methylation by the protein product of gene TK2122. (A) *The tertiary structure of rRNAs (CryoEM) where a loss in cytidine retention is detected (red, arrow) after RNA BS-seq of the strain deleted for gene TK2122.* ***(B, C)*** *The predicted structure of mRNAs that TK2122 methylates as demonstrated in vitro. The cytidine targeted for methylation is indicated by an arrow.* ***(D)*** *For each candidate m^{5}C site that was lost in the TK2122 deletion strain, the predicted secondary structure is aligned. "P" indicates paired and "S" indicates single stranded nucleotides. The cytidine targeted for methylation is at the center (position 21) and surrounded by 20 nt of up and downstream sequence.*



Structural analysis of RNAs targeted for methylation by the protein product of gene TK2304. (A-B) *The tertiary structure of rRNAs (CryoEM) where a loss in cytidine retention is detected (red, arrow) after RNA BS-seq of the strain deleted for gene TK2304.* ***(A)*** *C1957,* ***(B)*** *C2026, C2027 and C2047 in the 23S rRNA are either single or double stranded.* ***(C)*** *The predicted structure of the TK1911-mRNAs that TK2304 methylates as demonstrated in vitro. The cytidine targeted for methylation is indicated by an arrow.* ***(D)*** *For each candidate m^{5}C site that was lost in the TK2304 deletion strain, the predicted secondary structure is aligned. "P" indicates paired and "S" indicates single stranded nucleotides. The cytidine targeted for methylation is at the center (position 21) and surrounded by 20 nt of up and downstream sequence. Note*

that the predicted secondary structure of the rRNA is not entirely consistent with the Cryo-EM structure of the ribosome.

4) This reviewer does not see a reason why the authors would include the results showing bisulfite-refractory cytosines in 2 out of 3 repeats at all. Preferably, many more repeats should have been performed to arrive at robust data, which, as nicely shown by McIntyre et al., Scientific Reports (2020), increases the reproducibility of m6A detection in mRNA. If the data for 3/3 repeats exist, what is the use of excluding one repeat unless the authors try to prop up the numbers of potentially modified mRNAs from 77 to 232?

We agree that more replicates will usually always result in higher confidence in the data and its reproducibility, regardless of the nature of such data. We sequenced the control strain in triplicate and across 2 different conditions for a total of 6 experiments. We also sequenced both total RNA and rRNA-depleted pools, for a total of 12 sequencing libraries for our parent strain. For each deletion strain (n=17), we sequenced total RNA and rRNA-depleted RNA across two growth conditions and in duplicates, adding 136 more sequencing libraries. Based on our rigorous analysis protocols, artifact elimination measures, reproducibility and confidence thresholds, and *in vitro* orthogonal validation, we feel that the level of replication in this study is sufficient to address the research goals. We would also note that high-throughput sequencing experiments like ours are usually replicated 2-3 times, so our experimental design is not unusual and meets the standards accepted in the field. Although we appreciate that the reviewer is scrutinizing the accuracy of the data through, we respectfully decline to generate additional replicates.

5) Have the authors repeated RNA bisulfite sequencing of a particular locus that was predicted to be methylated by their high-throughput sequencing experiments? The perceived message of the manuscript is that only 77 mRNAs contained reproducible detectable cytosines after bisulfite treatment (3/3 experiments). However, the stoichiometry at a particular nucleotide seems to differ, averaging at 15-20%. Unless the authors designed the study including UMIs in the sequencing runs (the mentioning of which is nowhere to be found), it would be important to test for some of the positions in the 77 mRNAs, and ask if a locus is really bisulfite-refractory on more than a few mRNA molecules that were successfully reverse-transcribed into cDNA. This could strengthen or weaken the message by showing that the identified loci are indeed containing non-converted cytosines outside of deamination-resistant nucleotide sequences.

Since our goals were largely to identify candidate m^5C sites for mechanistic studies and identify the methyltransferases that install the m^5C epitranscriptome, we opted to

validate sites through *in vitro* methylation assays. To orthogonally verify a few m^5C sites identified by BS-seq, we performed *in vitro* methylation assays and demonstrated that the methyltransferases identified here site-specifically install m^5C at the cytidine shown to be bisulfite refractory *in vivo* (Figure 4D-G and Supp. Figure 9.IV).

Each library was barcoded, but we did not use UMIs. We describe in the materials and methods our methodology for eliminating PCR duplicates. The Samtools "rmdup" function identifies PCR duplicates by identifying pairs of reads where multiple reads align the 5' end of mate 1 and the 3' end of mate 2 to the same exact positions in the genome.

6) Sodium bisulfite treatment of RNA causes major degradation. Creating cDNA libraries from such RNAs with reduced base complexity (presence of AGU with very low levels of Cs remaining) will cause biases during cDNA synthesis. For instance, poly-U stretches due to deamination of cytosine within CUCU context might affect the efficiency of reverse transcriptase creating poly-A. In addition, poly-A stretches might cause PCR bias when amplifying cDNAs. The authors should explain how they determined that reads covering a specific genomic (transcript) region have been derived from many cDNA molecules and therefore from multiple bisulfite-treated RNAs. Or, they should state how they excluded mapped reads derived from a limited number of cDNA molecules that represent only a few "surviving" RNAs. This is especially important when looking at the non-converted positions in reads with high coverage. Since the authors do not mention barcoding of cDNA synthesis to address the potential for cDNA synthesis or PCR bias, performing this post-hoc exercise should allow the reader to gauge the extent of bias in the data sets.

We agree with the reviewer that it is important to process NGS data carefully and ensure that the proper steps are taken to reduce the impact of library construction bias in the mapped data. This is especially true in the context of lower complexity sequences derived from bisulfite treatment. In our case, we removed PCR duplicates using the following samtools commands:

"Using samtools, alignments were removed from bam files […] when detected as a PCR duplicate according to the Samtools manual using the following series of commands: *samtools view -h -b -q 20 <sampleID>_unsorted.bam | samtools sort -n -o <sampleID>_mapq_nsorted.bam; samtools fixmate -rm strain_mapq_nsorted.bam; <sampleID>_fixmate.bam; samtools sort <sampleID>_fixmate.bam > <sampleID>_fixmate_csorted.bam; samtools markdup -r <sampleID>_fixmate_csorted.bam <sampleID>_rmdup.bam*."

The Samtools “rmdup” function identifies PCR duplicates by identifying pairs of reads where multiple reads align the 5’ end of mate 1 and the 3’ end of mate 2 to the same exact positions in the genome. The materials and methods section has been modified to include this statement.

7) Related to the previous comment, the authors should provide information about the expression level of those mRNAs, which they flag as containing m5C. If those mRNAs are highly expressed, because they encode proliferation-relevant proteins and contain sequence features that mimic tRNAs or rRNAs, they might become substrates of some m5C circuitry and therefore appear to be methylated. In addition, any such mRNA identity that is represented by many individual molecules will “survive” the deamination reaction better than lowly expressed mRNA identity. This might uncover particular mRNAs, which appear to be more methylated just because they are highly expressed in the archaeon. The authors should try to address this possibility of skewing the read-out, which would then indicate or refute the existence of an artifact introduced by the applied methodology.

Supplementary File 1 records total coverage, m^5C coverage, whether the site is high confidence in 2 or 3 replicates, as well as an extensive annotation of each m^5C site.

8) The authors use CGmaptools to calculate C/T coverage. Does the software map cytosines in other contexts than CpG? And can they elaborate on why they did not use software that was designed to map RNA bisulfite sequencing data?

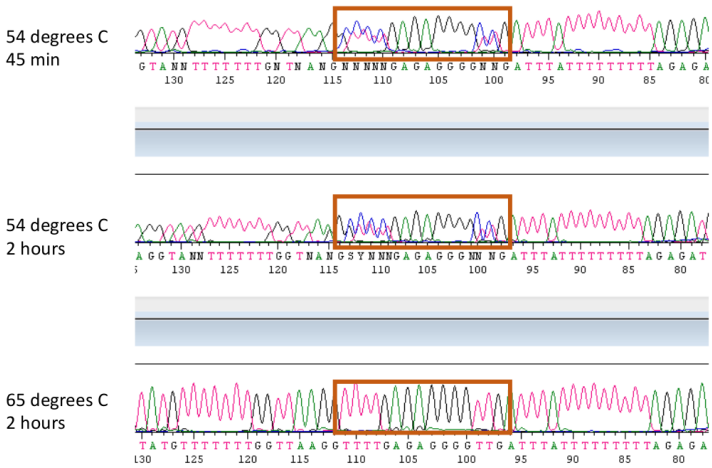
We used BSseeker 2 (which is a bowtie2 wrapper) to map reads to the reference genome. This is a mapping software that is specifically designed to map BS-seq reads. CGmaptools was used to calculate the coverage of each nucleotide (A, T, C, G) at each genomic position regardless of sequence context. We provide only the relevant CGmaps in a repository (please see data availability statement). All software used, custom and open source, was specifically designed to analyze bisulfite sequencing libraries. The Materials & Methods section details the software, software versions, and command line arguments used under the “Data processing” subsection.

9) In light of the small genome of *Thermococcus kodakarensis*, have the authors considered addressing the extent of bisulfite artifacts by transcribing the genome and using the “naked” RNA as a reference for their bisulfite sequencing? This approach has been described in pmid: 34594034.

Although this is a fascinating idea, we are concerned we would be unable to recapitulate the transcriptome in its mature form *in vitro*. Many genes in *Thermococcus* are transcribed in operons, generating immature forms of mRNA that would be differentially susceptible to bisulfite refractory artifacts. *In vitro* transcription of the genome would produce pre-rRNAs and pre-tRNAs that are similarly not reflective of the mature transcriptome. Since secondary structures result in increased number of cytidine retention (and BS-artifacts), BS-seq of the *in vitro* transcribed genome would not generate reliable artifact signals.

Bioinformatic efforts to remove “unconverted” reads were employed. In the rare instance when a sequenced read was composed of >3% cytidines, the read was removed from further analysis. Early efforts to establish BS-seq in *Thermococcus kodakarensis* involved determining the reaction conditions for cytidine deamination. We found that a 2 hour incubation at 65C resulted in complete deamination of a selected handful of *in vivo* transcribed RNAs.

Troubleshooting bisulfite conversion



Cytidines and thymines are represented by blue and pink traces, respectively.

10) Related to the sub-stoichiometry point, and given the low methylation levels for most of the identified m^5C sites in mRNA in eukaryotes (PMID: 31061524, 33330931, 34691665), it would be prudent to alert readers that even in the event of losing a particular m^5C site in the respective RNA modification enzyme knockout strain, the jury is still out there to prove beyond doubt that m5C at a particular mRNA position (in 1-2 out of 10 mRNAs, or for the “high coverage” mRNA, 5 out of 100) is biologically meaningful, especially under extreme conditions such as high temperatures.

A quick clarification, we employ several high confidence parameters, some of which are library specific. In order to address the potential differences in deamination rate in each library, we require that the m^5C coverage (not to be confused with total coverage) be at or above the 99th percentile for m^5C coverage across the transcriptome. For most libraries, this requires an m^5C coverage of ~6. At a 10% frequency threshold, that would equate to 6/60 (not 1-2/10). For sites where the m^5C coverage is at least 50x, we lower our threshold to 5%, meaning we allow sites where the m^5C coverage/total coverage is 50/1000 (not 5/100).

We agree that the biological impact of m^5C in mRNA coding sequences is unclear. We also acknowledge that even if a molecular consequence is proven (i.e. RNA stability, translation dynamics, etc), the fitness impact of mRNA m^5C is equally important to uncover. There are still many unknowns regarding the fate of RNAs once modified.

Specific comments:
Abstract: “While many modifications are limited in frequency, restricted to non-coding RNAs, or present only in select organisms, 5-methylcytidine (m5C) is abundant across diverse RNAs and fitness-relevant across Domains of life…"
> This reviewer asks to revisit the generalized statement. Currently, m5C is restricted to non-coding RNAs, and the few mRNA positions reported appear to be shrinking with improved technological repeat experiments. If one gives those data the benefit of the doubt, m5C in mRNA is about one magnitude lower than m6A, which is considered to be the most abundant with 1-3 modified Adenosines per average mRNA.
We are unsure of what the criticism here is. The statement is not saying that m^5C is restricted to non-coding RNAs, but that many other RNA modifications are while m^5C is present across diverse RNAs. This statement is not a contradiction of the reviewer's comment.

Page 4: “Modifications to even individual residues in long RNAs are known to elicit dramatic impacts on structure14–16, stability17–20, protein-interactions21, cellular…”
> Reference 15 and 16 refer to work on tRNAs, which are not long RNAs. Reference 21 is a review. Please, cite primary data.
References 15 and 16 have been removed. Reference 21 provides a nice review of proteins that interact with m^6A modifications, and we think this citation is appropriate and supports the statement. Additional primary references have been added.

Page 5: “Specialized nucleotides within rRNAs and tRNAs often provide essential stability and importantly, these modified RNAs must be passed to daughter cells, providing a heritable role for epitranscriptomic modifications.”

> It is unclear why modified RNAs must be passed to daughter cells. Please, provide primary references.

In single celled organisms, cell division results in two daughter cells, each receiving approximately half of the RNA from the parental cell. Regardless of how the RNA is divided between the daughter cells, rRNA and tRNAs are heavily and nearly stoichiometrically modified, and therefore daughter cells will inherently contain modified r- and tRNAs.

RNA modifications are abundant in haploid gametes and play a role in heritable phenotypes. Additional references have been incorporated into the manuscript to support that the epitranscriptome is heritable.

In mammals, sperm sRNAs are known to harbor various RNA modifications, and DNMT2 (m^5C tRNA MTase) is required for heritable m^5C modifications in sperm RNA. Deletion of cytidine deaminase (C→U editing) specifically results in higher risk of cancer in mice and in some cases, leads to embryonic lethality.

- Chen, Q. et al. Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science* 351, 397–400 (2016).
- Kiani, J. et al. RNA-mediated epigenetic heredity requires the cytosine methyltransferase Dnmt2. *PLoS Genet.* 9, e1003498 (2013).
- Nelson, V. R., Heaney, J. D., Tesar, P. J., Davidson, N. O. & Nadeau, J. H. Transgenerational epigenetic effects of the *Apobec1* cytidine deaminase deficiency on testicular germ cell tumor susceptibility and embryonic viability. *Proc. Natl Acad. Sci. USA* 109, E2766–E2773 (2012).

Eggs are also enriched for select RNA modifications. Fully grown mouse germinal vesicle oocytes and metaphase II eggs display abundant A-to-I editing in mRNAs compared to growing oocytes from postnatal day 12 oocytes. These inosines were enriched in mRNA protein coding regions (CDS) and specifically located at the third codon base, or wobble position.

- Pavla Brachova, Nehemiah S Alvarez, Xiaoman Hong, Sumedha Gunewardena, Kailey A Vincent, Keith E Latham, Lane K Christenson, Inosine RNA modifications are enriched at the codon wobble position in mouse oocytes and eggs, *Biology of Reproduction*, Volume 101, Issue 5, November 2019, Pages 938–949.

Page 5: “Evidence suggests this regulatory paradigm extends to mRNAs and across Domains, with RNA modifications being essential for all life”
> Most RNA modifications are not essential for life as shown in numerous mutagenesis screens in yeast and other organisms.
We do not mean to suggest that all individual sites of modification are essential to life. Individual modified residues are often not essential – we mean to say that the epitranscriptome as a whole is essential. We define the epitranscriptome as the totality of RNA modifications within a cell, inclusive of all ~170 known modification identities and their coordinate positions. For clarity, the statement has been modified to replace *“...RNA modifications being essential for all life*” with “*...with the epitranscriptome being essential for all life*”.

Despite the evolutionary conservation of many modification sites, many organisms do not exhibit growth phenotypes under laboratory conditions when lacking these modifications. But, many reports (including this report) do show that modifications support cell growth under stress conditions, and even individual modified residues have been shown to be essential under specific growth conditions. Here are a few examples relevant to our report:

- In *Thermus thermophilus*, a hyperthermophilic bacteria, tRNA 1-methyladenosie at position 58 (PMID: 12682365), 7-methylguanosine at position 46 (PMID: 19934251), *and* 2-Thioribothymidine at position 54 (PMID: 16317006) **are required for growth at high temperature**. In the absence of these tRNA modifications, *T. thermophilus* is viable at optimal growth conditions, but NOT viable under heat stress.
- In *Thermococcus kodakarensis*, tRNA 2-*dimethylguanosine at position 10 (PMID: 31405913) and 2’phosphouridine at position 47 (PMID: 35477761)* are viable under optimal growth conditions, but NOT viable under heat stress.

The following paragraph has been added to the introduction to clarify these points:

“Many archaea thrive in conditions inhospitable to most extant life. Maintaining RNA structure at high temperature, or at the extremes of salinity, pressure, or pH are facilitated, in part, by chemical modifications. Specific RNA modifications are critical for life at high temperatures, and loss of epitranscriptomic modifications is often lethal.[37–40] In *Thermus thermophilus*, a hyperthermophilic bacteria, tRNA 1-methyladenosie (m^1A)[39], 7-methylguanosine (m^7G)[41], *and* 2-Thioribothymidine[38] are required for growth at high temperatures. Likewise, *in Thermococcus kodakarensis, a hyperthermophilic archae*on, 2-dimethylguanosine (m^2_2G)[40] and 2’O-phosphouridine (p^2U) in tRNAs are required for growth at high temperatures. Not only are individual modified residues in tRNAs essential for hyperthermophilic growth, 4-acetylcytidine (ac4C) and 2’O-methyl-ac4C in *Pyrococcus furiosus* and ac4C in *T. kodakarensis* are markedly increased with rising growth temperature.[2,42] 5-methyl-2-thiouridine (m^5s^2U), m^2_2G, archaeosine (G+)

and m^1A stabilize *T. kodakarensis* tRNA structure and promote hyperthermophilic growth.[16,37,40]"

Page 5: "Site-specific, and often sub-stoichiometric, modification of mRNA is linked to core biological functions and responses to environmental changes."
> This statement cannot be a general one. Exactly the sub-stoichiometric modification is one of the great conundrums in RNA modification research. As of now, nobody has ever tested how many site-specific modifications are required for biological impact. Removing the writer enzymes in their entirety is insufficient to answer this question. Hence, stoichiometry of RNA modifications has not been linked to core biological functions.

We do not propose that the stoichiometry of individual RNA modifications has been linked to biological function. Rather, we state that RNA modifications, including those that are sub-stoichiometry, have been linked to core biological functions. We include citations to support this statement.

Page 6: "We identified at least 232 unique m^5C sites in a diverse set of coding and non-coding RNAs, established that ~10% of all unique transcripts contain m^5C, and that mRNA represents the largest fraction of m5C-modified RNAs…"
> This statement should read that the authors identified at least 77 unique sites refractory to bisulfite in 3 out of 3 experiments. Furthermore, that mRNA represents the largest fraction of m^5C-modified RNAs is only supported because rRNA and tRNA were depleted from the input RNA into the bisulfite sequencing. Hence, this statement is misleading and will therefore likely be picked up by AI-generated abstracts produced by paper mills, which will reiterate that the m^5C is only found in mRNA in this organism.

rRNAs were included in our analysis and we made numerous qualifications and stated the reason for why our libraries do not include tRNAs. We state that within the 232 sites that we detected m^5C, mRNAs constitute the largest fraction of unique transcripts that incorporate m^5C. This is a true and accurate statement based on the data, not our opinion. We felt it is appropriate and transparent to include the complete analysis of m^5C sites that were detected in 2/3 AND 3/3 replicates. We can not control whether artificial intelligence will consider our qualifications, but we are confident that human intelligence will understand we face technological challenges when sequencing tRNAs.

We employed a strict high confidence threshold to detect m^5C sites, leading to many m^5C sites falling slightly short in meeting this criteria in one replicate but meeting these strict threshold in the other 2 replicates. There were only a small few sites that were completely absent in one of the three replicates, but otherwise met our strict threshold in

the other 2 replicates. The biological conclusions are nearly identical when analyzing sites present in 2/3 or 3/3 replicates. For those who wish to only consider sites that meet our high confidence threshold in all three replicates, supplementary file 1 provides complete information about the confidence, coverage, etc of each site.

A good example is the m^5C site detected in TK1911. As described in the text, this site fell slightly below our high confidence threshold in one of three replicates under stationary growth conditions. We orthogonally validated that rTK2304 site-specifically installs m^5C in TK1911-mRNA *in vitro*. These data demonstrate that even though the m^5C site in TK1911 has low stoichiometry, this site can be modified in a site-specific manner by rTK2304, adding validity to the *in vivo* data.

Page 13: “Using samtools, alignments were removed from bam files where MAPQ score was < 20 and when detected as a PCR duplicate…”
> How did the authors identify PCR duplicates with that method?
The Samtools “rmdup” function identifies PCR duplicates by identifying pairs of reads where multiple reads align the 5’ end of mate 1 and the 3’ end of mate 2 to the same exact positions in the genome.

Page 16: “Visual inspection indicates that modification frequencies level-off at ~10%, indicating many modifications with a frequency below 10% may be false-positives. We determined that a high-confidence m^5C site must reach a 10% modification frequency.”
> What is visual inspection of high-throughput data? Why is a high confidence site determined by a 10% and not by a 90% modification frequency?
Please see Supplementary Figure 2D for a visual representation on high throughput data as described. A histogram of modification frequencies at each cytidine with at least 47X coverage shows that tens of thousands of bisulfite-refractory sites below a 10% “modification” frequency. A significant trend is observed that correlates modification frequency and number of sites at that frequency. In many cases, this trend is likely due to non-conversion artifacts or processive star activity of methyltransferases. At ~10% modification frequency (red vertical line), this trend is no longer apparent (red horizontal line). There are very likely legitimate m^5C sites below a 10% modification frequency, however, within these BS-seq experiments, we are not confident that we can distinguish a true positive from a false positive when modification frequencies are so low. For this reason, we applied a 10% minimum modification frequency threshold.

Nearly all BS-seq reports set a modification frequency threshold, although that threshold varies quite a bit study-to-study. It is common to see a 3-20% modification frequency threshold applied to call high-confidence sites, so our methodology is not out of line relative to standards set within the field.

T. kodakarensis is an extremophile and thrives in harsh conditions, and in our experience RNA derived from this species is more stable against spontaneous degradation. Concerns about incomplete bisulfite-conversion due to increased structural stability was an early concern. Although not included in this report, early efforts were dedicated to establishing the best bisulfite conversion reactions that lead to complete deamination of cytidines while maintaining RNA integrity.

We arrived to the final decision to perform the sodium bisulfite conversion reactions at 65°C for 2 hours. We achieved near complete conversion of sanger sequenced cDNA. In the BS-seq libraries reported here, we achieved an extremely high library-wide cytidine deamination rate (~99.8 to 99.9 %, Supplementary Figure 2B).

Page 21-22: “All 23 nt RNA substrates were ordered from IDT and resuspended in nuclease-free water”
> How did the authors arrive at this sequence length?
We reason that a 23 nt RNA would encode the necessary and sufficient sequence motifs required for site-specific modification. Relative to the small genome size of *T. kodakarensis* (~2 Mbp), approximately 11 nt are needed to encode a sequence that will uniquely target a protein to that sequence. Additionally, the molecular weight in kilodaltons of the methyltransferases investigated in this report are ~20-50 KDa. Based on these two premises, we reasoned that if the protein recognizes nucleotide sequence alone, 11 nt would likely encode the necessary and sufficient sequence motifs, and due to its size, the protein likely would only be large enough to bind 11 nucleotides. It is difficult to know if the protein binds the RNA upstream or downstream of the cytidine targeted for modification, so in addition to the cytidine targeted for modification, we included 11 nt up and down stream, resulting in a 23 nt substrate.

Page 25: “…hydroxymethylcytidine, 3-methyl cytidine, or 4-methyl cytidine, by comparing retentions times and mass transitions of corresponding modified nucleosides (data not shown).”
> Given that this organism is likely harboring more modifications than just m^5C, it would be commendable to show the detected nucleosides?
Although we understand the benefits to the scientific community to report all modifications we detected by LC-MS/MS as soon as possible, our thorough discussions on this matter led to the final decision to not report the totality of the mass spectrometry data. Our consortium is gearing up to publish several reports in addition to this report, including a comprehensive analysis of RNA modifications across *T. kodakarensis* and other archaeal, extremophilic species. In all honesty, the mass spectrometry data is quite extensive and the data presented here is already overwhelmingly large. It would

be a disservice to attempt to publish the BS-seq data and the mass spectrometry data in a single publication. Although the mass spectrometry data will be made publicly available, we politely decline to include the analysis of such data in this report, reserving it for an upcoming publication.

> The depiction of the mass spec data in the main Figure is insufficient. Normally mass spectra of chromatographic peaks with retention times are shown with m/z on the x-axis and the relative abundance of the modification on the y-axis (as provided in the supplementary data).

We thank the reviewer for their insightful comment. We have added additional information to each depiction of mass spec data in the manuscript. The nucleoside analysis presented in Figures 4F, S1B, S1C, S8C (left panel), and S8D (left panel) were collected utilizing a targeted approach on a triple quadrupole mass spectrometer. To detect a nucleoside of interest, we monitored a specific mass transition that is characteristic of this nucleotide. The retention time vs. intensity chromatograms shown in each of the Figures 4F, S1B, S1C, S8C, and S8D represent the mass response (intensity) relative to this transition measured at the given retention time. We have added the mass transitions which we monitored to each plot to provide a more complete depiction of these data.

The oligonucleotide analysis presented in Figures S8B, S8C (right panels) and S8D (right panels) correspond to deconvoluted mass vs. intensity spectra of a single chromatographic peak acquired by an orbitrap mass spectrometer. Figure S8A corresponds to a representative mass charge vs. intensity fragmentation spectra acquired by an orbitrap mass spectrometer. We have added the corresponding retention times for each chromatographic peak to provide a more complete depiction of these data.

Page 25: “RNA sequencing libraries were prepared for total RNA and mRNA-enriched fractions ...”
> It is unclear how the authors enriched for mRNA. All they did according to M & M is to deplete rRNAs or remove small RNAs (<200 nt) from total RNA.
The reviewer is correct. We did not enrich mRNA. All instances of mRNA-enrichment have been replaced with rRNA-depleted.

Page 26: “Owing to the depth of high-quality sequencing, such as that of sites with ≥ 1000x coverage, the acceptable minimum m5C frequency was lowered to 5%.”

> It is unclear why higher coverage should allow for including positions that are bisulfite-refractory in 5/100 transcripts.

It would be 50/1000 (m^5C coverage / total coverage) that would allow us to lower our modification frequency thresholds to 5%. Only at high coverage sites can we be confident in lower modification frequencies. Note we only achieved such coverage at a few sites, including the rRNA and TK0895, the S-layer protein. There are very few sites that actually meet this criteria. Supplementary Figure 1 illustrates the pain-staking efforts dedicated to evaluating the noise of each library and implementing dynamic parameters to call high confidence m^5C sites based on the conversion rates and noise present in each library.

The first implementation includes removing reads that had a >3% cytidine retention, indicating to us that these reads escaped the bisulfite conversion, perhaps due to retained secondary structures despite the harsh denaturing conditions. Visual inspection of these reads on a genome browser showed that these reads had horizontal tracts of many retained cytidines. These reads were easily distinguishable from other reads mapped to the same region, and usually accounted for a low percent (and low modification frequency) of reads. Since these reads are not representative of the vast majority of reads that mapped to the same region, they were eliminated from further analysis. This step alone massively improved our confidence in the data sets.

In most cases we demanded a 10% modification frequency at a 47X total coverage while the m^5C coverage minima is dynamic and depends on the library. In Supplementary Figure 2C, we illustrate a few examples in our control strain. At each high coverage (at least 47X) site with at least a 10% modification frequency, a histogram shows the number of sites at a range of m^5C coverages. The minimum m^5C coverage employed here is that of a 99th percentile. The minimum m^5C coverage usually fell at ~6x or a little higher in noisier libraries. That means in many cases, we required 6/60 $m^5C/(C+T)$ or greater.

For reference, we achieved > 5000X coverage of rRNA and nearly complete conversion of cytidines (that we did not detect as being modified). m^5C sites in rRNA usually have a very high modification frequency exceeding 75%, and the >20 m^5C sites detected in rRNA are among the highest confidence sites we detected. Apart from these 20 or so m^5C sites, the other cytidines were deaminated nearly completely, leaving 5-7 retained cytidines at > 5000X coverage. This is a modification rate of 0%. We know that there will always be some level of noise due to non-conversion events, whether these events are stochastic or due to secondary structures.

Page 29: “We mapped m^5C within GCG codons (alanine, n=18) to an extent that is fourfold higher than expected if by random chance. Modifications to codons GCU (alanine, n=13), GGC (glycine, n=18), CCG (proline, n=24), and CUG (leucine, n=16) were similarly enriched. When codons include multiple cytidines, there is a strong bias for which cytidine is selected for modification”
> High CG content has previously been connected to deamination artifacts. Could the authors provide secondary structure predictions of these sites within these mRNAs (similar to Figure 4H)? If these codons form secondary structures with neighboring sequences, then the creation of bisulfite artifacts is likely.
Structural analysis (predicted based on minimum free energy) shows no correlation with single or double stranded nucleotides. This was actually one of the first analyses we performed in hopes of identifying structural elements in substrate RNAs as well as potential non-conversion artifacts. Since we have not observed a correlation between single or double stranded nucleotides at m^5C sites, it is unlikely that secondary structure is having a large impact on bisulfite refractory artifacts, as far as we can tell.

Structural information at each candidate m^5C site is recorded in supplementary file 1, where “.” and “|” symbols represent predicted single and double stranded nucleotides, respectively. Each full length RNA was folded using viennaRNA fold (energy parameter modified to fold based on MFE at 85 degrees celsius), and the fold of 40 bases of the surrounding sequence is included.

Page 30: “Taken together, these data indicate a strong positional bias in m^5C sites in particular codon and amino acid contexts, likely indicating m^5C impacts the translatability of these codons.”
> m^5C does not affect base pairing but more stacking interactions. How would one m^5C-modified cytosine in 5 out of 100 mRNAs affect the translation of the encoded protein resulting in biological impact, especially under hyper-thermophile conditions?
This is a great question and one that remains to be answered. Whether/how these modifications drive fitness and impact the fate of mRNAs once modified are just beginning to be understood. This study provides a foundation to address these exact questions, and we are excited to pursue the molecular impact of high on low stoichiometric modifications on individual RNAs.

Page 32: “To our surprise, certain strains deleted for putative m^5C-specific and alternative RMTase enzymes showed gains in m^5C sites or increased modification frequencies.”
> What is an increased modification frequency? Please, provide these data to the reader. Unless these observations can be repeated with targeted RNA bisulfite sequencing and barcoded cDNA synthesis to control for RNA molecule input and,

hence number of cDNAs, this result could be interpreted as a sign of deamination artifacts, which are amplified by RNaseH (-) RT and PCR.
As stated in the text, we define increased (or decreased) modification frequency as at least a 2 fold increase in the cytidine retention rate at individual sites of bisulfite refractory. We reported relative increases/decreases in modification frequency in supplementary table 1. We understand that the modification frequencies we report may not be exact, but our primary goal is to look for absolute losses in m^5C sites and correlate each loss with the loss of a methyltransferase (shown in Table 2). Our focus is on absolute losses, but we also detected a few absolute gains and relative changes that we are reporting for the sake of data transparency.

Page 37: “Structural comparisons between the three RNA substrates (Figure 4H) adumbrate that the secondary and tertiary structures between these RNAs may play a role in the substrate recognition or catalytic activity of rTK2304 in vitro…”
> What is “adumbrate”?
According to Merriam-Webster dictionary; adumbrate is to foreshadow vaguely; to suggest, disclose, or outline partially.

Page 38: “To confirm modification of the TK1911 mRNA at the site predicted based on the loss of in vivo modification due to deletion of TK2304, the 101 nt substrate containing a C or U at the central position was mixed with unlabeled SAM and rTK2304 in vitro, purified, then digested to single nucleosides for LC-MS/MS analysis. We observed 0.37% of m5C/C ratio (or 5% oligo modification frequency) on the C-containing 101-nt RNA (Figure 4F and G)”
> Given that the enzyme TK2304, when acting in situ, might not methylate all substrates for various reasons, it remains unclear why a clean enzyme presented with a clean RNA is not able to methylate more than 5% of the position. If one takes any classic rRNA or tRNA methylating enzyme, those will modify pretty much every transcript at the target nucleotide position. The result of the in vitro methylation activity indicates that the methylation frequency is similar to the methylation background that was defined by the authors for the analysis of the bisulfite sequencing data. And, therefore, the results agree with the notion that this enzyme is likely not an enzyme acting on mRNA, hence cannot be called a robust enzyme (as stated by the authors in a downstream paragraph).
This is a great question and one that we have been pondering. There could be many reasons for this. rRNAs have much longer half-lives (hours) than mRNAs (seconds to minutes). It could be that quick turn over of mRNAs prevent their efficient methylation while rRNAs stick around for much longer and therefore are available to be modified at a higher rate. It is also possible that some modifications are required for efficient ribosome biogenesis and we are only sequencing RNA derived from mature rRNA.

Another possibility; If we think that RNA structures play a role in targeting methyltransferases to mRNAs to be methylated then it would reason that there may only be certain points during the life-cycle of the mRNA that would make that mRNA amenable to modification. There may be a short time co-transcriptionally that the RNA forms a particular structure that is methylated. Translation may be impacted by these co-transcriptional modifications, as transcription and translation are coupled in *T. kodakarensis*.

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
This reviewer stated before:

"The manuscript is well suited to contribute important information to the RNA modification community. Unfortunately, the authors chose to exclude to query both tRNA and rRNAs for m5C modification. Importantly, in light of the more recent findings that m5C in eukaryotic mRNA is catalyzed mostly by NSUN2 and NSUN6, two tRNA methyltransferases, and occurs largely at tRNA-like sequences (at much fewer mRNAs than previously thought; PMID: 31061524, 33330931, 34691665), an obvious question is if the few mRNA positions reported to be bisulfite conversion-refractory in the mRNA of this archaeon are representing technical artifacts, or are the consequence of Star activity of a particular m5C RNA modification circuitry, which evolved to modify rRNAs and tRNAs but also takes aim at similar sequence or structure in other RNAs. To approach an answer to the latter question, this reviewer asks the authors to implement some additional bioinformatics analysis, before publication of the manuscript in Nature Communications can be recommended. The additional analysis would help the field to generalize if prokaryotes, just like eukaryotes, have a propensity to allow mRNA modifications in sequence or structural context akin to non-coding RNAs."

The authors sent a long rebuttal letter, the contents of which are summarized as:
"Language has been added that clarify the function and fitness-relevance of m5C modifications to coding sequences may be a result of star activity of methyltransferase circuitry. Supplementary figure 1 has been expanded to include m/z mass spectrometry data at the request of reviewer 1. Additional experiments that measure the activity of methyltransferases on small RNA fractions have been performed at the request of reviewer 3. The introduction has been expanded and now includes a summary of what is known about hyperthermophilic epitranscriptomes. Additional bioinformatics analysis regarding RNA structure was performed at the request of reviewer 1 and 2. All grammatical and typographical errors were corrected.

Even though the authors opted for explaining away particular criticisms, and their explanations are a good read, the obviously abstained from performing particular experiments such as targeted bisulfite sequencing on candidate loci. This is a pity, and might be explained with the fact that Nature Communications has become a journal where investigators need to go after higher ranking journals rejected their work. Hence, doing more work to appease a referee has become a clear cost/benefit analysis. Well, while this reviewer found the addition of mass spectrometry data as requested, and the changes to the text, other requests were ignored and some of the described changes cannot be found, especially in regard to the stated additional bioinformatics analyses.

For instance, responding to a specific comment, the authors state: Structural information at each candidate m5C site is recorded in supplementary file 1, where "." and "|" symbols represent predicted single and double stranded nucleotides, respectively. Each full length RNA was folded using viennaRNA fold (energy parameter modified to fold based on MFE at 85 degrees celsius), and the fold of 40 bases of the surrounding sequence is included.
>> This reviewer was not able to find supplementary file 1, and therefore cannot judge if the authors have addressed the request and provided the data.
>> Supplementary file 1 has also been used in responses to other referees.

Furthermore, some of the responses of the authors need clarification.

(Previous) General comments:

1) It remains unclear why … potentially very low in this organism.

Author response:
"We apologize for the misunderstanding that we only analyzed mRNA. We analyzed RNA >200 nt by BS-seq and a variety of RNA size fractions by mass spectrometry (Supplementary Figure 1). We did not exclude rRNA. In fact, we included an extensive analysis of candidate m5C in rRNA. Figure 3 and Supplementary File 3 are entirely dedicated to m5C occurrence in the 16S-tRNA-23S rRNA operon…"
>> Supplementary Figure 3 is not dedicated to what the authors state.
Supplementary Figure 3. Supplementary Figure 3. Linear regression of m5C frequencies indicated reproducible modification frequencies.
>> Besides Figure 3, where is the data that contain an extensive analysis of rRNA?

(Previous) Specific comments:

7) Related to the previous comment, the authors should provide information about the expression level of those mRNAs, which they flag as containing m5C. If those mRNAs are highly expressed, because they encode proliferation-relevant proteins and contain sequence features that mimic tRNAs or rRNAs, they might become substrates of some m5C circuitry and therefore appear to be methylated. In addition, any such mRNA identity that is represented by many individual molecules will "survive" the deamination reaction better than lowly expressed mRNA identity. This might uncover particular mRNAs, which appear to be more methylated just because they are highly expressed in the archaeon. The authors should try to address this possibility of skewing the read-out, which would then indicate or refute the existence of an artifact introduced by the applied methodology.

Author response:
"Supplementary File 1 records total coverage, m5C coverage, whether the site is high confidence in 2 or 3 replicates, as well as an extensive annotation of each m5C site."

>> Again, where is supplementary File 1? However, this reviewer maintains that gene expression analysis from bisulfite-treated RNA is impossible since RNA degradation will introduce biases. Hence, the recorded total coverage as submitted in a non-existing supplementary file 1 cannot be taken as a data set to understand if the mRNAs containing potential m5C are highly or lowly expressed? There must be existing gene expression data on the organism within the consortium. This should be analyzed and included in the revised manuscript. This reviewer knows that highly expressed RNAs survive the bisulfite treatment quantitatively better than lowly expressed RNAs. Even though the authors repeatedly argue that the state-of-the art in the epitranscriptomics field is a particular number of repeats and controls, this reviewer insists on improving the quality of the data that is being published presently and I the future. Hence, the informed reader should know if the 77 or 232 mRNA identities are highly expressed in the organism.

Page 5: "Site-specific, and often sub-stoichiometric, modification of mRNA is linked to core biological functions and responses to environmental changes."

> This statement cannot be a general one. Exactly the sub-stoichiometric modification is one of the great conundrums in RNA modification research. As of now, nobody has ever tested how many site-specific modifications are required for biological impact. Removing the writer enzymes in their entirety is insufficient to answer this question. Hence, stoichiometry of RNA modifications has not been linked to core biological functions.

Author response:
We do not propose that the stoichiometry of individual RNA modifications has been linked to biological function. Rather, we state that RNA modifications, including those that are sub-stoichiometry, have been linked to core biological functions. We include citations to support this statement.
>> Reference 33-35 are papers, which confuse intergenerational with transgenerational effects that are caused by genetic aberrations (paramutation) or by dietary changes. While these papers pin all

intergenerational observations on RNAs and their modifications, these papers do not mention sub-stoichiometric numbers, and are therefore not the correct references for sub-stoichiometry when regarding modified RNAs.

Page 38: "To confirm modification of the TK1911 mRNA at the site predicted based on the loss of in vivo modification due to deletion of TK2304, the 101 nt substrate containing a C or U at the central position was mixed with unlabeled SAM and rTK2304 in vitro, purified, then digested to single nucleosides for LC-MS/MS analysis. We observed 0.37% of m5C/C ratio (or 5% oligo modification frequency) on the C- containing 101-nt RNA (Figure 4F and G)"

> Given that the enzyme TK2304, when acting in situ, might not methylate all substrates for various reasons, it remains unclear why a clean enzyme presented with a clean RNA is not able to methylate more than 5% of the position. If one takes any classic rRNA or tRNA methylating enzyme, those will modify pretty much every transcript at the target nucleotide position. The result of the in vitro methylation activity indicates that the methylation frequency is similar to the methylation background that was defined by the authors for the analysis of the bisulfite sequencing data. And, therefore, the results agree with the notion that this enzyme is likely not an enzyme acting on mRNA, hence cannot be called a robust enzyme (as stated by the authors in a downstream paragraph).
Author response:
"This is a great question and one that we have been pondering. There could be many reasons for this. rRNAs have much longer half-lives (hours) than mRNAs (seconds to minutes). It could be that quick turn over of mRNAs prevent their efficient methylation while rRNAs stick around for much longer and therefore are available to be modified at a higher rate. It is also possible that some modifications are required for efficient ribosome biogenesis and we are only sequencing RNA derived from mature rRNA. Another possibility; If we think that RNA structures play a role in targeting methyltransferases to mRNAs to be methylated then it would reason that there may only be certain points during the life-cycle of the mRNA that would make that mRNA amenable to modification. There may be a short time co-transcriptionally that the RNA forms a particular structure that is methylated. Translation may be impacted by these co-transcriptional modifications, as transcription and translation are coupled in T. kodakarensis.
>> This reviewer did not ask the authors to explain what could be the reason for the low in vitro activity of the enzyme, but was criticizing the statement of robustness. However, the authors still maintain the statement in the new version: "While methyltransferase activities for rTK2304, rTK1935, and rTK2122 were robust, specific, and validated by LC-MS/MS…"

In addition, this reviewer requests that the authors should include the data in the manuscript, which were used to answer general comment 3, even though the authors did not address the question as directly.

3) tRNAs and rRNAs are the most abundantly modified RNA species in any cell type…Can the authors compare the positions that are bisulfite-refractory in the few mRNAs with the sequences and structures of archaeal rRNAs and tRNAs (TKt41 and 16S-tRNAAla-23S)?

Author response:
"Below, we are providing 3 figures depicting our structural analysis of 3 bona fide RNA m5C methyltransferases reported here. We surveyed the available cryoEM structure of the T. kodakarensis ribosome for secondary and tertiary structures at m5C sites as well as secondary structure predictions using vienna RNAfold. We anticipated to see structural similarities at least between sites that we confirmed by mass spectrometry or from the cryoEM structure. However, we did not see large consistencies in RNA structures targeted by each enzyme."

>> Including these analyses as supplement would be very informative to the readers.

Reviewer #2:
Remarks to the Author:
The authors have adequately addressed my issues and those of the other reviewers, and have improved the manuscript. I do not have additional comments.

Reviewer #3:
Remarks to the Author:
Comments to the authors:

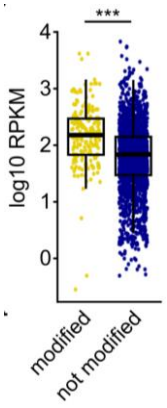
All my concerns have been addressed by the authors.
I believe that this is a nice work and recommend the editor to accept this manuscript.

REVIEWER COMMENTS

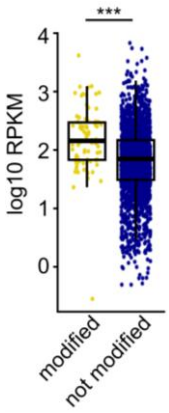
Reviewer #1 (Remarks to the Author):

>> Again, where is supplementary File 1? However, this reviewer maintains that gene expression analysis from bisulfite-treated RNA is impossible since RNA degradation will introduce biases. Hence, the recorded total coverage as submitted in a non-existing supplementary file 1 cannot be taken as a data set to understand if the mRNAs containing potential m5C are highly or lowly expressed? There must be existing gene expression data on the organism within the consortium. This should be analyzed and included in the revised manuscript. This reviewer knows that highly expressed RNAs survive the bisulfite treatment quantitatively better than lowly expressed RNAs. Even though the authors repeatedly argue that the state-of-the art in the epitranscriptomics field is a particular number of repeats and controls, this reviewer insists on improving the quality of the data that is being published presently and in the future. Hence, the informed reader should know if the 77 or 232 mRNA identities are highly expressed in the organism.

Based on their prior experience in this field, the reviewer suggests that modification frequency quantification is more accurate for highly expressed genes. The reviewer insists we perform gene expression analysis to assess whether m^5C-containing mRNAs are highly expressed or lowly expressed. As requested, we queried publicly available data from SRA for which RNA-seq was performed on cells grown under very similar laboratory conditions (datasets previously published by our consortium). This analysis showed that m^5C containing mRNAs are ~2-fold more abundant in RNA-seq datasets compared to mRNAs for which we did not map m^5C. Figure 1, Supplementary Figure 4, the main text [page12], and the materials and methods section have been updated to include this new data. We thank the reviewer for their recommendations and agree this additional analysis improves the manuscript.



(2/2 reps)



(3/3 reps)

Regarding bisulfite treatment influencing calculated modification frequencies: The exact numerical value of the modification frequencies may be confounded by the expression levels of

the transcript - as stated, highly expressed genes survive the harsh bisulfite treatment better than lowly expressed genes - but this is a reality of bisulfite sequencing. As it stands, bisulfite sequencing remains the *gold standard* for quantitative mapping of m^5C. RNA-seq datasets may be analyzed by zealously interested and expert readers, but there is not an obvious or intuitive way to correct modification frequencies with RNA expression levels, and it runs a very high risk of confusing readers. We maintain language describing modification frequencies as it was originally presented.

Regarding differential gene expression: We note that we did not perform differential gene expression analysis of the 16 strains from which we generated bisulfite-seq libraries, and there are no figures or text that intend to indicate such analyses were performed. Although several other studies have probed differential gene expression in the past, we have chosen to not investigate differential gene expression of m^5C-containing mRNAs or whole transcriptomes that lack select m^5C sites. In this study, we focus mainly on detecting m^5C sites and identifying the methyltransferases responsible for their installation. Although the bisulfite chemistry damages RNA and so our datasets cannot be used to measure differential or relative gene expression with confidence, we did use the coverage (not relative gene expression) as a confirmation that the RNAs were present (or sufficiently expressed) above a strict threshold that would allow modification frequency comparison across datasets (47x coverage as determined by a power calculation;see materials and methods). I will mention that this coverage/expression requirement far exceeds that of any other study I have read. We achieved exceedingly deep coverage owing to the high transcription levels and specific instrumentation used.

Regarding "Gene expression threshold" as described in the initial manuscript: Given the new gene expression analysis, we thought it best to update the text to distinguish the former language from the new text. To determine how many genes met our 47x expression threshold (as shown in Figure 1A), we simply asked how many genes met an average of 47x coverage across each nucleotide in each gene (Figure 1A). This is not totally accurate as I also include non-coding RNAs as "genes" in this explanation, but not in the manuscript. This was a necessary step to determine if our 47x coverage requirement for calling m^5C sites was too high; most genes had greater than 47x coverage across the gene – and most m^5C sites far exceeded 47x coverage (supplementary figure 2E) – so they are deemed sufficiently expressed in these datasets to allow a 47x coverage requirement for high confidence m^5C sites. This is not to say whether genes are highly or lowly expressed, just that they are in fact expressed. We indicate in the text and Figure 1A that ~10% of sufficiently expressed transcripts contain at least 1 m^5C site. The mass spectrometry data (supplementary figure 1) actually suggests many more m^5C residues likely exist than what we have been able to confidently call from the bisulfite-seq data, likely because the sequencing data are subject to many strict thresholds and confidence parameters to ensure the elimination of most false-positives (supplementary figure 2), which the mass spectrometry data is not subjected to. We have modified the language throughout the manuscript to more accurately and clearly reflect these sentiments. Mainly, we more clearly define and use the words "sufficiently expressed". The "Unique transcript expression threshold" section in the materials and methods has also been substantially edited.

[Abstract]
"...the m^5C epitranscriptome includes ~10% of unique transcripts."
-> "...the m^5C epitranscriptome includes ~10% of unique transcripts **sufficiently expressed in these data**."

"...established that ~10% of all unique transcripts contain m^5C, and that mRNA represents the largest fraction of m^5C-modified RNAs."
→ "...established that ~10% of all unique **and sufficiently expressed** transcripts contain m^5C, and that mRNA represents the largest fraction of m^5C-modified RNAs."

"*T. kodakarensis* encodes just 2,306 open reading frames and ~1900 unique transcripts (~82%) were expressed at or above our detection threshold."
→ " *T. kodakarensis* encodes just 2,306 open reading frames and ~1900 unique transcripts (~82%) were expressed at or above our detection threshold **(>47x average coverage across the gene)**."

[Materials & Methods, "Unique transcript expression threshold" subsection] Now reads,
→ "To determine the proportion of unique transcripts with and without modification as represented in Figure 1A, we calculated the number of unique transcripts that met an average of 47x coverage across all nucleotides in each unique transcript (including genes and non-coding RNAs). As such, unique transcripts with an average coverage ≥ 47x at each nucleotide were considered to be expressed at or above our expression thresholds. The expression threshold was applied in ≥ 2 or 3 replicates for m^5C sites reproducible in ≥ 2 replicates or 3 replicates (Supplementary Figure 4A). Most unique transcripts had > 47x average coverage – and most m^5C sites exceeded 47x coverage (supplementary figure 2E) – so they are deemed sufficiently expressed in these datasets to allow a 47x coverage requirement for identifying high confidence m^5C sites. We did not perform differential gene expression analysis as bisulfite treatment degrades RNA and results in sequencing datasets biased for longer RNAs, and therefore can not be quantitatively reliable."

> Page 5: "Site-specific, and often sub-stoichiometric, modification of mRNA is linked to core biological functions and responses to environmental changes." This statement cannot be a general one. Exactly the sub-stoichiometric modification is one of the great conundrums in RNA modification research. As of now, nobody has ever tested how many site-specific modifications are required for biological impact. Removing the writer enzymes in their entirety is insufficient to answer this question. Hence, stoichiometry of RNA modifications has not been linked to core biological functions.

Author response:

We do not propose that the stoichiometry of individual RNA modifications has been linked to biological function. Rather, we state that RNA modifications, including those that are sub-stoichiometry, have been linked to core biological functions. We include citations to support this statement.

>> Reference 33-35 are papers, which confuse intergenerational with transgenerational effects that are caused by genetic aberrations (paramutation) or by dietary changes. While these papers pin all intergenerational observations on RNAs and their modifications, these papers do not mention sub-stoichiometric numbers, and are therefore not the correct references for sub-stoichiometry when regarding modified RNAs.

Author response:
To err on the side of caution, we have removed the word “substoichiometric” and edited this sentence to instead read, “Site-specific modification of mRNA is linked to core biological functions and responses to environmental changes.”

>Page 38: “To confirm modification of the TK1911 mRNA at the site predicted based on the loss of in vivo modification due to deletion of TK2304, the 101 nt substrate containing a C or U at the central position was mixed with unlabeled SAM and rTK2304 in vitro, purified, then digested to single nucleosides for LC-MS/MS analysis. We observed 0.37% of m5C/C ratio (or 5% oligo modification frequency) on the C- containing 101-nt RNA (Figure 4F and G)”. Given that the enzyme TK2304, when acting in situ, might not methylate all substrates for various reasons, it remains unclear why a clean enzyme presented with a clean RNA is not able to methylate more than 5% of the position. If one takes any classic rRNA or tRNA methylating enzyme, those will modify pretty much every transcript at the target nucleotide position. The result of the in vitro methylation activity indicates that the methylation frequency is similar to the methylation background that was defined by the authors for the analysis of the bisulfite sequencing data. And, therefore, the results agree with the notion that this enzyme is likely not an enzyme acting on mRNA, hence cannot be called a robust enzyme (as stated by the authors in a downstream paragraph).

Author response:
“This is a great question and one that we have been pondering. There could be many reasons for this. rRNAs have much longer half-lives (hours) than mRNAs (seconds to minutes). It could be that quick turn over of mRNAs prevent their efficient methylation while rRNAs stick around for much longer and therefore are available to be modified at a higher rate. It is also possible that some modifications are required for efficient ribosome biogenesis and we are only sequencing RNA derived from mature rRNA. Another possibility; If we think that RNA structures play a role in targeting methyltransferases to mRNAs to be methylated then it would reason that there may only be certain points during the life-cycle of the mRNA that would make that mRNA amenable to modification. There may be a short time co-transcriptionally that the RNA forms a particular structure that is methylated. Translation may be impacted by these co-transcriptional modifications, as transcription and translation are coupled in *T. kodakarensis*.

>> This reviewer did not ask the authors to explain what could be the reason for the low in vitro activity of the enzyme, but was criticizing the statement of robustness. However, the authors still maintain the statement in the new version: “While methyltransferase activities for rTK2304, rTK1935, and rTK2122 were robust, specific, and validated by LC-MS/MS…”

Author response:
We apologize for misunderstanding the original comment. The reviewer’s point has become more clear. The sentence has been modified to exclude the word “robust” and now instead reads, “While methyltransferase activities for rTK2304, rTK1935, and rTK2122 were specific and validated by LC-MS/MS …”. We agree this statement is now more accurate.

>>In addition, this reviewer requests that the authors should include the data in the manuscript, which were used to answer general comment 3, even though the authors did not address the question as directly. *3) >tRNAs and rRNAs are the most abundantly modified RNA species in any cell type…Can the authors compare the positions that are bisulfite-refractory in the few mRNAs with the sequences and structures of archaeal rRNAs and tRNAs (TKt41 and 16S-tRNAAla-23S)?*

Author response:
We were initially hesitant to include structural analysis of mRNAs, but as the reviewer finds these data helpful, we have included them. Supplementary figures 10, 11 and 12 have been added. We thank the reviewer for this recommendation.

>> “…they obviously abstained from performing particular experiments such as targeted bisulfite sequencing on candidate loci….”

Author response:
We appreciate the reviewer's meticulous assessment of this manuscript and continuous efforts to ensure its rigor and quality. We appreciate the exchange of ideas that ultimately improve the manuscript.

Our method of orthogonal validation, which we feel is not only appropriate but superior to targeted bisulfite sequencing, is aimed at validating bisulfite-refractory sites while also identifying the *bona fide* enzymes that generate m^5C at those sites. We validated 5 m^5C sites in mRNA, and identified the enzyme that installs m^5C at those sites. Based on the available CryoEM structure of the *Thermococcus kodakarensis* ribosome, we found densities consistent with m^5C at 12 sites identified in the bisulfite-seq datasets. Several cytidine residues we identified as candidate m^5C sites were not resolved in the structure and could not be validated.

17 total m^5C sites were validated using *in vitro* experimentation or available CryoEM data.

To summarize - We took a bisulfite-seq approach to identify m^5C sites in parent/control strain and strains deleted for methyltransferases. This *in vivo* data correlated losses in candidate m^5C (bisulfite-refractory) sites with the loss of individual methyltransferases. This data strongly

suggested that the deleted methyltransferase is directly responsible for installing m^5C at the site where the bisulfite-refractory signal was lost. These *in vivo* data were then orthogonally validated using *in vitro* biochemical assays which confirmed these sites were occupied by *bona fide* m^5C modifications. We feel that our *in vitro* assays that include purified components and LC-MS/MS demonstrate the site-specific targeting activity of several methyltransferases on several mRNA targets, and this data is highly validating; these m^5C sites not only exist, but are installed by enzymes with high specificity. The *in vivo* data in combination with the *in vitro* data are in agreement and exceptionally convincing of substoichiometric m^5C in mRNA.

As the reviewer previously pointed out – although methyltransferases may reproducibly and site-specifically target these sites for m^5C installation, these mRNAs may not be the evolutionarily intended targets. Further research is required to address whether m^5C sites in mRNA are “off-target” or have any fitness benefits that drive evolution.

There are several ways (with slight variations) to perform targeted bisulfite sequencing as reported in the literature:

1. Use DNA probes to capture mRNAs of interest, purify them away from total RNA, bisulfite-treat and sequence these isolated RNAs.
2. Bisulfite treat total RNA then purify RNAs of interest for sequencing
3. Bisulfite treat total RNA then perform cDNA synthesis only of RNAs of interest

We understand that other groups have performed targeted bisulfite sequencing using the methods listed above to validate bisulfite-refractory sites. However, I feel this approach is most useful when researchers aim to validate bisulfite-refractory sites identified by other groups or previous experiments with the intention of following up on those previous experiments.

In our case, we are having trouble understanding how more bisulfite sequencing can act as an orthogonal approach to validate the bisulfite sequencing we performed within the same study. In this study, we have generated > 200 bisulfite sequencing datasets and analyzed them in parallel. It is highly unlikely that additional bisulfite-seq experiments will impact the biological conclusions or change the existing data in any way. We will also add, a cost/benefit analysis is relevant, as the process from start to finish (growing up biomass, purifying RNA, bisulfite treating, sequencing, and analysis) is expensive, laborious, computationally intensive, requires substantial memory/storage quotas, and is extremely time consuming. However, the cost/benefit of executing experiments was not a main consideration. We are dedicated to publishing only high quality, rigorously acquired, and orthogonally validated research.

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
This reviewer is thanking the authors for their repeated efforts to improve the manuscript, data provision and data presentation as well as the detailed rebuttal letter to the second installment of peer review.
The manuscript has improved to the extent that it can be accepted for publication in Nature Communication provided that the authors
a) amend a number of statements that are prone to misinterpretation
b) include some missing references where statements are not supported, and
c) provide a link to the source of the publicly available data for the RNA seq analysis resulting in Figure 1D and Supplementary Figure 4D.

Specific comments:
Line 71-73: "We demonstrate that m5C is ~25x more abundant in T. kodakarensis than human cells and the m5C epitranscriptome includes ~10% of unique transcripts sufficiently expressed in these data."
>> Transcripts are not expressed in data. Data is information allowing us to infer which unique transcripts were sufficiently expressed in the organism. Please, re-phrase.

Line 268: "…retrieved RNA-seq data from NCBI Sequence Read Archive generated from cultures grown…"
>> Please, provide the source for this statement.

Line 271: "It is likely the m5C sites mapped to mRNA in this study are better identified due to the higher expression levels of these mRNAs, whereas modifications to lowly expressed mRNAs may be more confounded."
>> This reviewer still maintains that the authors have not exactly understood what was meant by her/his insistence to include gene expression data allowing to factor in the relative mRNA expression levels of those potential RNA methyltransferase substrates that survive the harsh chemical treatment using sodium bisulfite. This conclusion is warranted because the authors still phrase the interpretation of the data in a biased manner. The bias lies in the assumption that sequences containing bisulfite-refractory cytosines must have contained modified nucleotides. This bias is not only represented by the statement on line 271 but also in the rebuttal letter concluding that: "The reviewer insists we perform gene expression analysis to assess whether m5C-containing mRNAs are highly expressed or lowly expressed."

It follows that this reviewer has been misunderstood since her/his question was not "…whether m5C-containing mRNAs are highly expressed or lowly expressed." but was based on the knowledge that highly expressed RNAs survive the bisulfite treatment quantitatively better than lowly expressed RNAs. With an open mind and keeping the considerable chance of RNA deamination artifacts in mind, this reviewer pointed at the possibility that any RNA existing in multiple copy numbers (highly expressed) will have a higher chance (than lowly expressed RNAs) to be recovered after the destruction of RNA sequence by sodium bisulfite. Hence, the presented finding is a confirmation of the notion that RNAs with bisulfite-refractory cytosines (due to deamination artifact or not) appear to be higher expressed in a given cell or organism. This is so because any RNA that is lowly expressed will have simply disappeared from the RNA pool and cannot be reverse-transcribed. However, the authors turn this argument now into a statement reading that m5C-modified mRNAs can be better identified since they are highly expressed. Which could be true but cannot be stated while the potential for interpreting artifacts as m5C is still valid. If the authors would have used UMIs at cDNA synthesis to approximate the number of RNAs represented by cDNAs and PCR products, one would be better positioned to argue about artifacts or real m5C sites. In light of finding particular bisulfite-refractory frequencies at specific sites, this reviewer would like to remind the authors that 40-1000x coverage

can be achieved easily from a few surviving RNA sequences after bisulfite treatment, which still could be the amplification of 20% of molecules containing a deamination artifact. This reviewer would like to keep an open mind, and have the readers remain vigilant, too, especially in relation to potential for artifacts that could be interpreted as m5C marks on mRNA. Please, re-phrase in such a way that the uninformed reader is not concluding that highly expressed mRNAs contain the most m5C marks.

Line 361-362: "We identified multiple sites where an absolute loss in m5C modification frequency (≤ 2%) was observed in one of five deletion strains."
>> Is this percentage correct? Say, the stated modification frequencies in mRNA presented in that study range from 5-65%. Would a total loss in m5C translate into something from 5% to 3% and 65% to 63%? Please, confirm or re-state.

Line 391-392: "This suggests that T. kodakarensis, and perhaps other species, may generate epitranscriptomes with overlapping and complex activities for distinct RNA modifications."
>> "…and perhaps other species…" Try to abstain from speculation in the Results section.

Line 419: "Ribosomes across bacterial and eukaryotic model species typically harbor 2-3 m5C sites."
>> Ribosomal RNA not ribosomes harbor m5C sites.

Line 578-581: "tRNAs were not captured well in our sequencing libraries despite repeated attempts, and therefore a comprehensive analysis of tRNA m5C profiles was not performed."
>> This statement sounds like the authors attempted to sequence tRNA. However, tRNA was excluded from the RNA preps as the statement in the next sentence justifies: "Although tRNA and other small RNAs were largely removed from sequencing libraries…". Please, re-phrase.

Line 581-586: tRNAs or tRNA-containing sequences?
>> Please, do not call tRNA sequence-containing reads tRNAs, especially if they have extended lengths. Mature tRNAs have a length between 70 and 90 nucleotides. Anything longer is not going to fit into the ribosome and therefore must be a precursor or a transcript that does not function as tRNA. Please, abstain from calling those potentially modified RNAs tRNAs and re-phrase.

Line 635-636: "…we have shown that mRNAs dominate the pool of m5C-containing RNAs in T. kodakarensis."
>> This is statement is misleading. Considering the whole "RNA pool" of a cell, ribosomal and tRNA make up > 95% of the total RNA mass. mRNA mass is ≤ 3%. Considering the sequence identity of RNAs, rRNAs and tRNAs are limited in unique sequences while mRNA is presenting the most diverse RNA species. If one would consider RNA mass carrying m5C, rRNA and tRNA will be dominating, if one considers unique sequences, mRNA might be dominating. Please, re-phrase so readers do not understand that mRNA mass is more modified than non-coding RNAs.

Line 670: "Previous work has indicated that some m5C sites in eukaryal models increase the stability of mRNAs."
>> Please, cite appropriate references for this statement to hold.

Line 682: "Several reports have demonstrated that deletion of genes encoding R5CMTs leads to hypersensitivity to heat stress in rice and worms.70,68."
>> Two, not several reports have demonstrated this notion. Please, do not overstate.

Line 684: "The selective employment of m5C over m6A, and its unprecedented abundance leads us to speculate that m5C may increase the thermal stability of RNAs."
>> What is the basis for this musing that m5C is as versatile as m6A as an RNA modification? Is there a reason why one would/should expect m6A to be in a system or not? In light of the author's refusal to include the identification of other RNA modifications by LC/MS, this reviewer can only assume that

m6A is rather absent in this organism. However, the uninformed reader does not know that. Hence, this statement come out of nowhere and is confusing as such.

Line 719-720: “It has long been speculated but sparse evidence would indicate whether R5CMTs share partial redundancy in target sites.”
>> Please, cite appropriate references for this statement to hold.

Response to Reviewer #1:

This reviewer is thanking the authors for their repeated efforts to improve the manuscript, data provision and data presentation as well as the detailed rebuttal letter to the second installment of peer review.
The manuscript has improved to the extent that it can be accepted for publication in Nature Communication provided that the authors
a) amend a number of statements that are prone to misinterpretation
Several statements have been amended. See marked text and information below for details.

b) include some missing references where statements are not supported, and
Additional references were added. See marked text and information below for details.

c) provide a link to the source of the publicly available data for the RNA seq analysis resulting in Figure 1D and Supplementary Figure 4D.
The materials and methods section includes the SRA accession numbers for data presented in these panels. The data availability statement has been updated to also include the accession numbers per the reviewer's request.

Specific comments:
Line 71-73: "We demonstrate that m5C is ~25x more abundant in T. kodakarensis than human cells and the m5C epitranscriptome includes ~10% of unique transcripts sufficiently expressed in these data."
>> Transcripts are not expressed in data. Data is information allowing us to infer which unique transcripts were sufficiently expressed in the organism. Please, re-phrase.
"...in these data" has been removed.

Line 268: "…retrieved RNA-seq data from NCBI Sequence Read Archive generated from cultures grown…"
>> Please, provide the source for this statement.
Citation added.

Line 271: "It is likely the m5C sites mapped to mRNA in this study are better identified due to the higher expression levels of these mRNAs, whereas modifications to lowly expressed mRNAs may be more confounded."
>> This reviewer still maintains that the authors have not exactly understood what was meant by her/his insistence to include gene expression data allowing to factor in the relative mRNA expression levels of those potential RNA methyltransferase substrates that survive the harsh chemical treatment using sodium bisulfite. This conclusion is warranted because the authors still phrase the interpretation of the data in a biased manner. The bias lies in the assumption that sequences containing bisulfite-refractory cytosines must have contained modified nucleotides. This bias is not only represented by the statement on line 271 but also in the rebuttal letter concluding that: "The reviewer insists we perform gene expression analysis to assess whether m5C-containing mRNAs are highly expressed or lowly expressed."

It follows that this reviewer has been misunderstood since her/his question was not "…whether m5C-containing mRNAs are highly expressed or lowly expressed." but was based on the knowledge that highly expressed RNAs survive the bisulfite treatment quantitatively better than lowly expressed RNAs. With an open mind and keeping the considerable chance of RNA deamination artifacts in mind, this reviewer pointed at the possibility that any RNA existing in multiple copy numbers (highly expressed) will have a higher chance (than lowly expressed RNAs) to be recovered after the destruction of RNA sequence by sodium bisulfite. Hence, the presented finding is a confirmation of the notion that RNAs with bisulfite-refractory cytosines (due to deamination artifact or not) appear to be higher expressed in a given cell or organism. This is so because any RNA that is lowly expressed will have simply disappeared from the RNA pool and cannot be reverse-transcribed. However, the authors turn this argument now into a statement reading that m5C-modified mRNAs can be better identified since they are highly expressed. Which could be true but cannot be stated while the potential for interpreting artifacts as m5C is still valid. If the authors would have used UMIs at cDNA synthesis to approximate the number of RNAs represented by cDNAs and PCR products, one would be better positioned to argue about artifacts or real m5C sites. In light of finding particular bisulfite-refractory frequencies at specific sites, this reviewer would like to remind the authors that 40-1000x coverage can be achieved easily from a few surviving RNA sequences after bisulfite treatment, which still could be the amplification of 20% of molecules containing a deamination artifact. This reviewer would like to keep an open mind, and have the readers remain vigilant, too, especially in relation to potential for artifacts that could be interpreted as m5C marks on mRNA. Please, re-phrase in such a way that the uninformed reader is not concluding that highly expressed mRNAs contain the most m5C marks.

The manuscript has been updated to explicitly state that m^5C is not necessarily enriched in highly expressed transcripts. The statement has been updated to include "...This is not to say m^5C is enriched in highly expressed transcripts, rather it is likely cytidines resistant to bisulfite deamination (m^5C or artifactual m^5C) are better identified due to a higher abundance of these RNAs (and therefore better able to survive harsh bisulfite treatment), whereas m^5C in lowly expressed RNAs may not be detectable at all."

Line 361-362: "We identified multiple sites where an absolute loss in m5C modification frequency (≤ 2%) was observed in one of five deletion strains."

>> Is this percentage correct? Say, the stated modification frequencies in mRNA presented in that study range from 5-65%. Would a total loss in m5C translate into something from 5% to 3% and 65% to 63%? Please, confirm or re-state.

A total loss would include bisulfite refractory sites where a high confidence and reproducible m^5C site (≥ 5% or 10% depending on abundance) in the parent strain is reduced to ≤ 2%.

Neither 5% → 3% or 65% → 63% constitute a total loss. Rather, 5% → 2% or 65% → 2% would

constitute a total loss. Changes in modification frequencies at sites of absolute m^5C loss are illustrated in Figures 4C and Supplementary Figure 9 A-D.III.

As a single example, C2062 in the 23S rRNA (supp file 3) is modified in 76-87% of ribosomes. When TK1935 is deleted, this modification frequency is reduced to 0%. In a strain doubly deleted for TK1935 and TK2304, the modification frequency is reduced to 1%. In both cases, this is considered a total loss in m^5C, and therefore the protein product of gene TK1935 is predicted to install m^5C at C2062 in the 23S rRNA.

Line 391-392: “This suggests that T. kodakarensis, and perhaps other species, may generate epitranscriptomes with overlapping and complex activities for distinct RNA modifications.”
>> “…and perhaps other species…” Try to abstain from speculation in the Results section.

The statement has been modified to read, “This suggests that *T. kodakarensis* ~~, and perhaps other species,~~ may generate epitranscriptomes with overlapping and complex activities for distinct RNA modifications.”

Line 419: “Ribosomes across bacterial and eukaryotic model species typically harbor 2-3 m5C sites.”
>> Ribosomal RNA not ribosomes harbor m5C sites.
The statement now reads, “Ribosomal RNA across bacterial and eukaryotic model species…”

Line 578-581: “tRNAs were not captured well in our sequencing libraries despite repeated attempts, and therefore a comprehensive analysis of tRNA m5C profiles was not performed.”
>> This statement sounds like the authors attempted to sequence tRNA. However, tRNA was excluded from the RNA preps as the statement in the next sentence justifies: “Although tRNA and other small RNAs were largely removed from sequencing libraries…”. Please, re-phrase.

Early attempts to sequence tRNAs did indeed fail. We have rephrased the statement to read, “tRNAs were not captured well in our sequencing libraries despite repeated attempts, and therefore a comprehensive analysis of tRNA m5C profiles was not performed.”

Line 581-586: tRNAs or tRNA-containing sequences?
>> Please, do not call tRNA sequence-containing reads tRNAs, especially if they have extended lengths. Mature tRNAs have a length between 70 and 90 nucleotides. Anything longer is not going to fit into the ribosome and therefore must be a precursor or a transcript that does not function as tRNA. Please, abstain from calling those potentially modified RNAs tRNAs and re-phrase.

We have carefully reworded the paragraph to make this distinction.
The statement has been added:

→ “The genomic regions that encode TKt41 and TKt30 are 140 nt and 169 nt long, respectively.”

Two sentences have been modified:

→ "....transcripts with TKt41 ($tRNA^{Trp}$) and TKt30 ($tRNA^{Ile}$) sequences persisted, likely in their immature form when their transcript lengths were longer…"
→ "Transcript positions corresponding to C113 in $tRNA^{Trp}$ and C51 in $tRNA^{Ile}$ are modified across growth conditions…"

Line 635-636: "…we have shown that mRNAs dominate the pool of m5C-containing RNAs in T. kodakarensis."
>> This is statement is misleading. Considering the whole "RNA pool" of a cell, ribosomal and tRNA make up **>** 95% of the total RNA mass. mRNA mass is ≤ 3%. Considering the sequence identity of RNAs, rRNAs and tRNAs are limited in unique sequences while mRNA is presenting the most diverse RNA species. If one would consider RNA mass carrying m5C, rRNA and tRNA will be dominating, if one considers unique sequences, mRNA might be dominating. Please, re-phrase so readers do not understand that mRNA mass is more modified than non-coding RNAs.
The statement has been corrected and now reads, "we have shown that mRNAs dominate the pool of unique, m5C-containing RNAs in *T. kodakarensis*."

Line 670: "Previous work has indicated that some m5C sites in eukaryal models increase the stability of mRNAs."
>> Please, cite appropriate references for this statement to hold.

Citation has been added.

Line 682: "Several reports have demonstrated that deletion of genes encoding R5CMTs leads to hypersensitivity to heat stress in rice and worms.70,68."
>> Two, not several reports have demonstrated this notion. Please, do not overstate.

The word "several" has been substituted for "two".

Line 684: "The selective employment of m5C over m6A, and its unprecedented abundance leads us to speculate that m5C may increase the thermal stability of RNAs."
>> What is the basis for this musing that m5C is as versatile as m6A as an RNA modification? Is there a reason why one would/should expect m6A to be in a system or not? In light of the author's refusal to include the identification of other RNA modifications by LC/MS, this reviewer can only assume that m6A is rather absent in this organism. However, the uninformed reader does not know that. Hence, this statement come out of nowhere and is confusing as such.

An analysis of m^6A in *T. kodakarensis* is included in Supplementary Figure 1B and discussed in paragraph two of the results section. We show that m^6A is not absent, but its abundance is very low compared to the universal human reference RNA. We have modified this sentence to

include a figure reference. The sentence now reads, “The selective employment of m^5C over m^6A (Supplementary Figure 1B), and its unprecedented abundance leads us to speculate….”

Line 719-720: “It has long been speculated but sparse evidence would indicate whether R5CMTs share partial redundancy in target sites.”
>> Please, cite appropriate references for this statement to hold.

It is difficult to provide a reference for a lack of evidence; We will not provide a reference where evidence of overlapping substrate specificity has not been demonstrated. Just as the reviewer has reflected on their own experiences regarding bias introduced by the degradation effects of bisulfite treatment, we reflect on our experiences interacting with the epitranscriptomics community. Speculation of overlapping substrates has been discussed within the community. To better reflect our experiences, we have slightly modified the statement to read, “It has been speculated but sparse evidence would indicate whether methyltransferases share partial redundancy in target sites.”

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
All remaining points raised by this reviewer have been addressed. This reviewer recommends publication and would like to thank the authors for engaging in the process of responding to all critical comments, for including suggestions by the reviewer, and for clarifying the scientific message of their work.