Perspectives on topology of the human m¹A methylome at single nucleotide resolution

XUSHEN XIONG,^{1,2,5} XIAOYU LI,^{1,5} KUN WANG,¹ and CHENGQI YI^{1,3,4}

¹State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, China

²Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China

³Department of Chemical Biology and Synthetic and Functional Biomolecules Center, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

⁴Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China

ABSTRACT

N¹-methyladenosine was recently reported to be a chemical modification in mRNA. However, while we identified hundreds of m¹A sites in the human transcriptome in a previous work, others have detected only nine sites in cytosolic and mitochondrial mRNAs. Herein, we provide additional evidence that hundreds of m¹A sites are present in the human transcriptome. Moreover, we show that both the improper bioinformatic tools and the poor quality of sequencing data in a previous study led to the failure in identifying the majority of m¹A sites. Our analysis hence provides an explanation of the divergence in the prevalence of this newly discovered mRNA mark.

INTRODUCTION

N¹-methyladenosine (m¹A) is a prevalent modification in stable noncoding RNAs including tRNA and rRNA. In 2016, m¹A was also found to be present in mammalian mRNA, with an m¹A/A ratio of ~0.02% by quantitative mass spectrometry (Dominissini et al. 2016; Li et al. 2016). Using a commercially available m¹A antibody, m¹A was found to enrich within 5' UTR of mRNA transcripts (Dominissini et al. 2016; Li et al. 2016). More recently, we and the Schwartz laboratory have independently developed single-nucleotide resolution m¹A methods, based on m¹A-induced misincorporation signature during reverse transcription (RT) (Li et al. 2017; Safra et al. 2017). Reassuringly, both studies found that (i) a known tRNA modification complex TRMT6/61A also works on mRNA; (ii) m¹A is present in the mitochondrial-encoded mRNA transcripts and dynamically responds to environmental stress or developmental cue; and (iii) m¹A within the coding sequence of mRNA can interfere with translation. However, while we identified 418 m¹A sites and recapitulated a 5' UTR enrichment pattern in the cytosolic mRNA, Safra et al. (2017) reported only nine sites in mRNA. Now, Dr. Schwartz has reanalyzed the data from the two studies and claimed that all of the 5'-UTR m¹A sites are misidentified due to

improper bioinformatics analysis ("duplications, misannotations, mismapping, SNPs, sequencing errors, and a set of sites originating from the very first transcribed base" [Schwartz 2018]). By analyzing the experimental data and comparing the bioinformatics algorithms, we now provide additional evidence for these sites. Moreover, we also provide potential explanations for the failure of the Schwartz study to identify these m¹A sites.

RESULTS AND DISCUSSION

Schwartz (2018) first converted the transcriptomic coordinates to genomic coordinates to allow comparison between data sets. During the conversion process, he found 82 redundant entries, which belong to transcripts with distinct RefSeq IDs but are mapped to the same genomic locus. We agree that these sites can be redundant. Yet, when using the genomic coordinates, we additionally detect 56 and 18 m¹A sites in the "promoter" region (i.e., within a 200 nucleotide [nt] region upstream of the annotated TSS) and introns, respectively, hence further expanding the m¹A methylome in nuclear-encoded mRNA (Supplemental Table S1). Schwartz also noted three sites that did not form part of the current RefSeq database

⁵These authors contributed equally to this work.

Corresponding author: chengqi.yi@pku.edu.cn

Article is online at <http://www.rnajournal.org/cgi/doi/10.1261/rna.067694.118>.

© 2018 Xiong et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

FIGURE 1. (Legend on next page)

and 37 sites that did not harbor an A at the reported position. Based on the 37 sites that he failed to convert, he calculated a false conversion rate of ~7.8% and estimated a further ~39 of the remaining sites to be false positive. However, we have now readily converted all 40 sites to the genomic position using the proper version of the annotation file (RefSeq, Release 71, see [Supplemental Table S2](#)). Hence, the aforementioned false positive rate does not hold water.

Schwartz (2018) also reclassified 196 5'-UTR m¹A sites to be exactly at the first nucleotide of transcripts. First of all, being a TSS site (whether annotated by RefSeq or "reclassified" by Schwartz) is independent of the modification status of a site. In fact, we have already suggested previously that a subset of these sites could potentially be TSS sites. Second, the evidence presented by Schwartz is not sufficient to conclude that all the 5'-UTR m¹A sites should be "reclassified" as TSS sites. For instance, a recent analysis has shown that a sizeable fraction of CAGE tags (~24% for an average gene) actually do not represent bona fide TSS sites (Adiconis et al. 2018). In addition, Schwartz used CAGE data from A549 cells, but not HEK293 cells used in the two studies; yet it is known that TSS usage may be cell-type dependent. To further evaluate the sufficiency of TSS site reclassification by terminal misincorporation, we analyzed the misincorporation pattern for the known m¹A modification at position 9 in both cytosolic and mitochondrial tRNA, which are also close to the 5' end of RNA. We found that the majority of the misincorporations are localized at the first two bases of the reads (Fig. 1A). A further meta-analysis of all the m¹A9 modifications in mt-tRNA shows a similar pattern to that of the "reclassified TSS" m¹A sites by Schwartz (Fig. 1B; compare to Figure 1C in Schwartz's manuscript). As a comparison, the m¹A sites in mt-rRNA, which are not clustered at the 5' end but are relatively evenly distributed throughout the transcripts (similar to the TRMT6/61A sites in mRNA), demonstrate a pattern of evenly spread mutation signals across the sequencing reads (Fig. 1B). Hence, the evidence provided by Schwartz is not sufficient to conclude that all the 5'-UTR m¹A sites are TSS sites. Future experiments are need-

ed to allow precise annotation of their position (TSS sites or 5'-UTR sites) in a transcript. Nevertheless, regardless of the position information discussed above, site annotation does not affect modification prevalence.

Schwartz (2018) further proposed that the "reclassified" TSS sites are computationally misidentified sites caused by nontemplated nucleotide incorporation during RT. First, we were aware that the nontemplated nucleotide incorporations may lead to mutation at the first base of a sequencing read, and we now plot the mismatch pattern of our data so as to formulate the influence of the nontemplated additions on m¹A calling. As shown in Figure 1C, for reads aligned to both tRNA and mRNA, the first base of a read indeed demonstrates a higher mutation rate compared to its following positions, showing mutational noise caused by nontemplated additions. However, such mutations caused by nontemplated additions do not occur exclusively nor preferentially to adenosines; they occur to C/G/T as well. If m¹A calling were based solely on the mismatch rate, a genuine m¹A-induced mismatch could indeed be difficult to distinguish from the background noise. However, modification identification in m¹A-MAP is based on the "difference" of the mismatch rate between the "IP" and "IP + demethylase" samples. Mutation caused by nontemplated additions should not be affected by demethylase treatment while true m¹A signals will be sensitive to demethylation. Hence, we plotted the mutation rate for the first base of all sequencing reads along mRNA transcripts for both the "IP" and "IP + demethylase" samples. A clear decrease of the mismatch rate was observed for A within the first ~200 nt of the transcripts (Fig. 1D), which nicely matches the enrichment of m¹A sites within this region (the red line in Fig. 1E). As a control, no reduction of mismatch rate was observed for the T/C/G bases in the transcriptome reference (Fig. 1D). In addition, we identified far less T/C/G sites under the same bioinformatics criteria (versus hundreds of A sites; Fig. 1E), strongly arguing against the claim that the "reclassified" TSS m¹A sites are artifacts due to nontemplated additions (the flat [hence no enrichment] lines in Fig. 1E). This observation is in sharp contrast to the previous study that uses RNA-DNA

FIGURE 1. Reanalyses of sequencing data by Li et al. (2017) and Safra et al. (2017) provide clear evidence of the presence of m¹A in mRNA. (A) IGV views of the known m¹A site at position 9 of a cytosolic and mitochondrial tRNA, showing enriched misincorporations at the first two positions of sequencing reads. (B) Barplots depicting the distribution of misincorporation events along the first 50 positions of the sequencing reads. Misincorporation in m¹A9 sites (in blue) are highly skewed to occur in the first two positions, whereas misincorporation in mt-rRNA sites (in red) are relatively uniformly distributed. (C) Mismatch rates of the first five positions in sequencing reads of tRNA (left panel) and mRNA (right panel) in the input sample. (D) The mismatch rate of the first base in all reads of the "IP" and "IP + demethylation" samples, plotted along the mRNA transcripts. Upon demethylase treatment, a clear decrease of mutation rate is specifically observed for A but not T/C/G. Bin size is 10 nt. (E) Distribution of the A/C/G/T sites that pass the detection threshold by Li et al. (2017). Bin size is 10 nt. (F) A "reverse calculation" to evaluate the potential false positives caused by m¹A-independent mutations. Mutation rates in the "IP" and "IP + demethylation" samples are artificially exchanged and modification calling procedure is performed using the same criteria. Very limited sites can be detected in such reverse calculation. (G) A known m¹A site at position 9 in an isoform of tRNA^{Asp(GUC)} can be identified by the end-to-end aligning mode, but not the soft clipping mode. Data taken from Li et al. (2017). (H) IGV view of an m¹A site at the 5' end of the ANKRD13A transcript. While the m¹A-induced mismatch rate is decreased upon demethylase treatment ("A" site within the black lines), mutation likely arising from nontemplated additions remains the same ("C" site 5' to the m¹A site).

difference for editing identification, which is also used as an evidence by Schwartz (Li et al. 2011); while this study reported similar numbers of editing events to all four bases,

our analysis above retrieved predominantly A sites. To further demonstrate the robustness of the mismatch rate difference-dependent m¹A calling, we performed a “reverse

FIGURE 2. (Legend on next page)

calculation" in which we artificially took the mismatch rate in the "IP + demethylase" samples as signal and that in IP samples as background. In addition to retrieving negligible T/C/G sites, we identified very limited A sites as well (see all four flat lines in Fig. 1F). Collectively, these analyses strongly argue against the claim that the m¹A sites were identified via "explicitly filtering." Instead, we demonstrate above that m¹A sites are identified in an unbiased manner throughout the transcriptome, based on the difference of mismatch rate that can clearly distinguish genuine m¹A-induced mutational signature from the background noise.

Schwartz (2018) also mentioned the use of STAR aligner that uses soft clipping mode for reads mapping and trims reads from the ends to maximize the alignment score. However, under the situations in which cDNA synthesis (and hence sequencing reads) terminates at the m¹A-induced, misincorporated nucleotide, the m¹A-induced mutational signals (at the first base of reads) could be simultaneously lost in the soft clipping mode. To better explain the difference between soft clipping mode and the end-to-end mode used by Li et al. (2017), we use an isoform of tRNA^{Asp(GUC)} as a demonstration (left panel in Fig. 1A). This isoform carries a known m¹A modification at position 9 (Cozen et al. 2015; Zheng et al. 2015; Clark et al. 2016), which was detected by Li et al. (2017) but missed by Safra et al. (2017). When utilizing the end-to-end mode, we find a high and low mutation rate in the IP and IP + demethylase sample, respectively, giving rise to a large difference of mutation rates that is used to identify m¹A (left panel of Fig. 1A). In contrast, the soft clipping mode removes the vast majority of mutation signals at this position, leading to the loss of a genuine m¹A site (Fig. 1G). The same situation also applies to mRNA, particularly for m¹A sites that are located at the 5' end of transcripts. One example is shown in Figure 1H, which is a cap + 1 m¹A site and was independently validated (Li et al. 2017). Additional examples can be found in the reanalysis by Schwartz: His Figure 1B shows three m¹A sites whose mismatch rates are decreased upon demethylase treatment, while his Figure 3 shows three negative sites whose mismatch rates are unaffected by demethylase treatment.

Does the difference in computational strategy fully explain the discrepancy between the two studies? To under-

stand the underlying causes, we also compared the data quality of the two studies. For the eight TRMT6/61A sites detected by both studies, a medium of ~1200 coverage was seen in the IP samples by Li et al. (2017) compared to less than dozens of reads by Safra and colleagues (Fig. 2A). As for the TRMT6/61A and TSS sites that were missed by Safra et al. (2017), a median coverage of ~50–100 was observed by Li et al. (2017) compared to less than 10 reads per site by Safra and colleagues (Fig. 2B). Then, what could explain the huge difference in useful reads (Fig. 2A,B), despite the fact that the raw sequencing reads by Safra and colleagues and Li et al. (2017) are actually comparable (within twofold)? We find severe reads duplication and rRNA contamination in the data of Safra et al. (2017) (Fig. 2C), possibly explaining why their data barely covered the sites identified in Li et al. (2017). Between the Dimroth reaction used by Safra and colleagues and enzymatic demethylation used by Li and colleagues, we previously performed a side-by-side comparison and already showed that the Dimroth reaction causes significant RNA degradation (see Figure S2E in Li et al. 2017).

What is even more surprising is the observed low reaction efficiency in the study by Safra et al. (2017), which significantly decreases the sensitivity of m¹A detection. For the two known m¹A sites in 16S and 28S rRNAs, a ~65% and ~63% mismatch rate still remains after the Dimroth reaction in the data of Safra et al. (2017), respectively, compared to a remaining mutation of ~1% after AlkB treatment (Fig. 2D) for both sites. Similarly, for m¹A575 in 12S mt-rRNA, which is identified *de novo* and biochemically validated by Li and coworkers (see Figs. 4C and S5B in Li et al. 2017) but missed by Safra et al. (2017), both their signal (IP sample) and the validation (IP + demethylation sample) data are of limited quality (Fig. 2E,F). Considering that m¹A sites in abundant rRNA were even missed, it is anticipated that detection of m¹A sites in low abundance mRNA would be very difficult in the Safra et al. (2017) study. More differences in experimental procedures that lead to the different qualities of the sequencing data sets by the two studies have been described in this technology preview (Dominissini and Rechavi 2017).

Schwartz (2018) also mentioned that they found "no significant change in stop rates" in their SSIII data and hence

FIGURE 2. Reanalyses of sequencing data by Safra et al. (2017) provide an explanation as to why they failed to identify the majority of m¹A in mRNA. (A) Reads' coverage for the eight TRMT6/61A sites (detected in both studies) in different sequencing data sets. The *inset* box in each panel provides a zoomed-in view of reads' coverage by Safra and coworkers. (B) Reads' coverage for the 44 TRMT6/61A sites and 196 "reclassified" TSS sites in different sequencing data sets. All these sites are missed in the study by Safra and coworkers. The *inset* box in each panel provides a zoomed-in view of reads' coverage by Safra et al. (2017). (C) Reads' duplication levels and rRNA contamination levels of different sequencing data sets. (D) Mismatch rates in the "IP" and "IP + demethylation" samples for m¹A1322 of 28S rRNA and m¹A947 of 16S mt-rRNA. (E) Mismatch rate in the "IP" and "IP + demethylation" samples for a novel m¹A575 site of 12S mt-rRNA. This site is biochemically validated by Li et al. (2017) but missed by Safra et al. (2017). (F) IGV views for the raw sequencing data and misincorporations for m¹A575 site of 12S mt-rRNA. (G) Reanalysis of the SSIII data by Safra et al. (2017) showing poor data quality that does not even support their own TGIRT data. The analysis was performed by aligning the reads to the genome reference using TopHat2. (H) IGV views for one modification site that is claimed as false positive due to its coincidence with SNP by Schwartz (2018). (I) IGV views for one modification site that is claimed as false positive due to its location within a polyC stretch.

questioned the TSS m¹A sites. Unfortunately, the coverage for the TSS sites in their SSIII data is extremely low (a medium of ~1–2) (Fig. 2B). More surprisingly, by reanalyzing the data, we found that their SSIII data is even inconsistent with their own TGIRT data: For instance, 4/10 sites have no truncation in IP samples while one site has no change of truncation rate after the Dimroth reaction (Fig. 2G). Only two sites show a decreased truncation rate, while the remaining three have no coverage at all. Hence, the SSIII data cannot even be used to support the m¹A sites reported by their own paper (Fig. 2G; Safra et al. 2017). Collectively, we do not think the quality of the data is sufficient to allow further analysis.

Schwartz (2018) used (i) a weak sequence motif and (ii) a very lax structural constraint to define TRMT6/61A-dependent m¹A sites, which lacks experimental evidence (Safra et al. 2017). As a matter of fact, the crystal structure of a TRMT6/61A-tRNA complex disfavors such loose requirements (Finer-Moore et al. 2015). During the reclassification by Schwartz, even sites fulfilling only one of these two criteria are considered TRMT6/TRMT61A substrates. This obviously leads to artificial inflation of the number of TRMT6/61A-dependent m¹A sites.

Schwartz (2018) speculated four sites overlapping with known SNPs and six sites within polyC stretches to be misidentified m¹A. We want to emphasize once again that m¹A identification is based on difference of mismatch rate between two experimental conditions where SNP will not show any difference. In fact, we showed that m¹A-MAP is capable of discriminating true m¹A sites from SNP (Fig. 2H, also see Figure 2D,E in Li et al. 2017). Likewise, mismatch rate difference-dependent modification calling should discriminate true m¹A sites from polyC stretches-induced sequencing errors as well (Fig. 2I). Except for finding coincidence of m¹A sites for SNPs and polyC stretches, Schwartz presented no evidence to support the strong claim.

Schwartz (2018) claimed ultra-low stoichiometry of m¹A sites based on the mutation rate in the input samples. Because of the limited sequencing depth for individual sites in transcriptome-wide data, we would recommend transcript-specific experiments that are very cost-effective and can simultaneously provide ultra-deep sequencing coverage, as we have done previously (see Figure 3D in Li et al. 2017, which has several thousand folds of coverage). The modification level as proxied by the mutation rate of targeted sites can then be more reliably estimated.

In conclusion, both optimized experimental procedures and tailored computational approaches are critical during the development of new sequencing tools. We anticipate that advanced epitranscriptomic technologies will continue to lead to exciting discoveries.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

REFERENCES

- Adiconis X, Haber AL, Simmons SK, Levy Moonshine A, Ji Z, Busby MA, Shi X, Jacques J, Lancaster MA, Pan JQ, et al. 2018. Comprehensive comparative analysis of 5'-end RNA-sequencing methods. *Nat Methods* **15**: 505–511.
- Clark WC, Evans ME, Dominissini D, Zheng G, Pan T. 2016. tRNA base methylation identification and quantification via high-throughput sequencing. *RNA* **22**: 1771–1784.
- Cozen AE, Quartley E, Holmes AD, Hrabeta-Robinson E, Phizicky EM, Lowe TM. 2015. ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nat Methods* **12**: 879–884.
- Dominissini D, Rechavi G. 2017. Loud and clear epitranscriptomic m¹A signals: now in single-base resolution. *Mol Cell* **68**: 825–826.
- Dominissini D, Nachtergaele S, Moshitch-Moshkovitz S, Peer E, Kol N, Ben-Haim MS, Dai Q, Di Segni A, Salmon-Divon M, Clark WC, et al. 2016. The dynamic N¹-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530**: 441–446.
- Finer-Moore J, Czudnochowski N, O'Connell JD III, Wang AL, Stroud RM. 2015. Crystal structure of the human tRNA m¹A58 methyltransferase-tRNA₃^{Lys} complex: refolding of substrate tRNA allows access to the methylation target. *J Mol Biol* **427**: 3862–3876.
- Li M, Wang IX, Li Y, Bruzel A, Richards AL, Toung JM, Cheung VG. 2011. Widespread RNA and DNA sequence differences in the human transcriptome. *Science* **333**: 53–58.
- Li X, Xiong X, Wang K, Wang L, Shu X, Ma S, Yi C. 2016. Transcriptome-wide mapping reveals reversible and dynamic N¹-methyladenosine methylome. *Nat Chem Biol* **12**: 311–316.
- Li X, Xiong X, Zhang M, Wang K, Chen Y, Zhou J, Mao Y, Lv J, Yi D, Chen XW, et al. 2017. Base-resolution mapping reveals distinct m¹A methylome in nuclear- and mitochondrial-encoded transcripts. *Mol Cell* **68**: 993–1005.
- Safra M, Sas-Chen A, Nir R, Winkler R, Nachshon A, Bar-Yaacov D, Erlacher M, Rossmann W, Stern-Ginossar N, Schwartz S. 2017. The m¹A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* **551**: 251–255.
- Schwartz S. 2018. M1A within cytoplasmic mRNAs at single nucleotide resolution: a reconciled transcriptome-wide map. *RNA* (this issue). doi: 10.1101/ma.067348.118.
- Zheng G, Qin Y, Clark WC, Dai Q, Yi C, He C, Lambowitz AM, Pan T. 2015. Efficient and quantitative high-throughput tRNA sequencing. *Nat Methods* **12**: 835–837.

RNA

A PUBLICATION OF THE RNA SOCIETY

Perspectives on topology of the human m¹A methylome at single nucleotide resolution

Xushen Xiong, Xiaoyu Li, Kun Wang, et al.

RNA 2018 24: 1437-1442 originally published online August 21, 2018
Access the most recent version at doi:[10.1261/rna.067694.118](https://doi.org/10.1261/rna.067694.118)

Supplemental Material

<http://rnajournal.cshlp.org/content/suppl/2018/08/21/rna.067694.118.DC1>

Related Content

m¹A within cytoplasmic mRNAs at single nucleotide resolution: a reconciled transcriptome-wide map
Schrage Schwartz
[RNA November , 2018 24: 1427-1436](#)

References

This article cites 12 articles, 2 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/24/11/1437.full.html#ref-list-1>

Articles cited in:
<http://rnajournal.cshlp.org/content/24/11/1437.full.html#related-urls>

Creative Commons License

This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>