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...
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. (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 + demethylase” 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 + demethylase” 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^metOCUt 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

![Image of Figure 2](https://example.com/figure2.png)

**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 tRNAAsp(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 re-analysis 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 understand 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 mtrRNA, 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...
questioned the TSS m\(^1\)A sites. Unfortunately, the coverage for the TSS sites in their SSIII data is extremely low (a medium of \(\sim 1-2\)) (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 m\(^1\)A 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) speculated four sites overlapping with known SNPs and six sites within polyC stretches to be mis-identified m\(^1\)A. We want to emphasize once again that m\(^1\)A known SNPs and six sites within polyC stretches to be mis-

The m\(^1\)A methylome in eu-

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 m\(^1\)A signals: now in single-base resolution. Mol Cell 68: 825–826.
Dominissini D, Natcheragaele S, Moshitch-Moshkovitz S, Peer E, Ko N, Ben-Haim MS, Dai Q, Di Segni A, Salmon-Divon M, Clark WC, et al. 2016. The dynamic N\(^1\)-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 m\(^1\)A58 methyltransferase-tRNA\(^{16S}\) 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, Young 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. Transcrito- wide mapping reveals reversible and dynamic N\(^1\)-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 m\(^1\)A methylome in nuclear- and mitochondrial-encoded transcripts. Mol Cell 68: 993–1005.
Safa M, Sas-Chen A, Nir R, Winkler R, Nachshon A, Bar-Yaacov D, Erlacher M, Rossmanith W, Stern-Ginossar N, Schwartz S. 2017. The m\(^1\)A landscape on cytosolic and mitochondrial mRNA at single-base resolution. Nature 551: 251–255.
Schwartz S. 2018. M1A within cytoplasmic mRNAs at single nucleo-
tide 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 sequenc-
ing. Nat Methods 12: 835–837.