PhOxi-Seq: Single-nucleotide resolution sequencing of N^2-methylation at guanosine in RNA by photoredox catalysis

Kimberley Chung Kim Chung, Yasaman Mahdavi-Amiri, Christopher Korfmann, and Ryan Hili*

Department of Chemistry and Centre for Research on Biomolecular Interactions, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada.

ABSTRACT: Chemical modifications regulate the fate and function of cellular RNAs. Newly developed sequencing methods have allowed a deeper understanding of the biological role of RNA modifications; however, the vast majority of post-transcriptional modifications lack a well-defined sequencing method. Here, we report a photo-oxidative sequencing (PhOxi-seq) approach for guanosine N^2-methylation — a common methylation mark seen in N^2-methylguanosine (m^2G) and N^2,N^2-dimethylguanosine ($m^2{}_2G$). Using visible light-mediated organic photoredox catalysis, m^2G and $m^2{}_2G$ are chemoselectively oxidized in the presence of canonical RNA nucleosides, which results in a strong mutation signature observed during sequencing. PhOxi-seq was demonstrated on various tRNAs and rRNA to reveal N^2-methylation with excellent response and markedly improved read-through at $m^2{}_2G$ sites.

Post-transcriptional modifications can have profound effects on the function and processing of RNA.[1] Not surprisingly, their dysregulation has been implicated in numerous human diseases,[2] especially various types of cancer.[3] With over 160 known RNA modifications identified across all life,[4] deciphering their biological roles remains a significant unmet challenge. Sequencing technologies that can resolve RNA modifications at single-nucleotide resolution have been instrumental to functional studies. Indeed, methods that have mapped the locations of N^6-methyladenosine (m^6A), 5-methylcytidine (m^5C), inosine (I), pseudouridine (Ψ), N^1-methyladenosine (m^1A) and 5-hydroxylmethylcytidine (hm^5C) have provided deeper insights into their impact on RNA function and metabolism,[5] and found application to disease diagnostics, prognostics, and epigenetic therapy.[6] Nevertheless, most post-transcriptional modifications lack a defined sequencing approach, which has hampered their functional elucidation.

N^2-methylguanosine (m^2G) and N^2,N^2-dimethylguanosine ($m^2{}_2G$) are prevalent modifications in RNA (Figure 1a). While mostly believed to be involved in structural tuning of RNAs,[7] recent reports have begun to shed light on the biological significance of N^2-methylation of guanosine. Studies have shown that upregulated m^2G in sperm transfer RNA-derived small RNAs (tsRNA) is an epigenetic factor that mediates intergenerational inheritance of metabolic diseases.[8,9] Furthermore, TRMT11, an m^2G writer, has been implicated in poor prognosis in prostate cancer, while mutations in TRMT1, an $m^2{}_2G$ writer, have been tied to aberrant human brain development[10] and autosomal-recessive intellectual disabilities.[11] Developing robust functional studies on N^2-methylation of guanosine has been limited by the absence of sequencing methods. While $m^2{}_2G$ blocks Watson-Crick-Franklin base-pairing and can often be detected during sequencing due to its high error signature,[12] only computational predictors exist for m^2G,[13,14,15] with no direct sequencing method available. The development of a robust single-nucleotide resolution sequencing method for N^2-methylation of guanosine in RNAs is expected to enable a better understanding of these modifications in human health and disease.

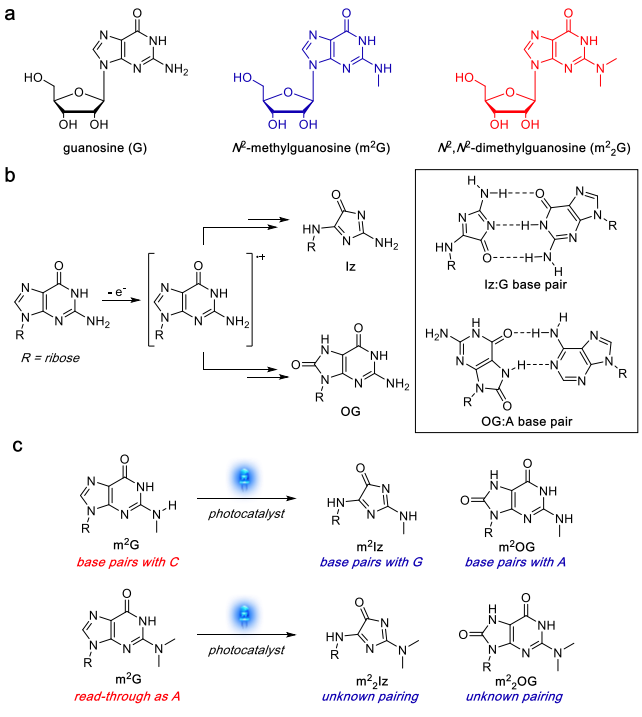

Figure 1. a) Guanosine and its N^2-methylated derivatives. b) oxidation pathway of guanosine into mutagenic derivatives along with their base-pairing partners (inset). c) Potential products formed during the photooxidation of N^2-methylated guanosine derivatives.

Guanosine is the most readily oxidizable nucleoside; its photooxidation in nucleic acids is known to occur by single electron oxidation to generate the guanine radical cation (Figure 1b).[16] This species can degrade *via* two oxidative pathways, and the outcome is governed largely by the nucleic acid environment.[17] In single-stranded or double-stranded forms, guanine is oxidized to 2,5-diamino-4H-imidazol-4-one (Iz), while in G-quadruplex form, guanine is preferentially oxidized to 8-oxoguanine (OG). Both Iz and OG are mutagenic forms of guanine. OG is known to base-pair with adenine, resulting in a G→T transversion, while Iz is known to base-pair with guanine, resulting in a G→C transversion (Figure 1b).[18] With these oxidation outcomes in mind and inspired by examples of visible light photoredox chemistry on RNA[19-20] and DNA,[21] we sought to explore the prospect of visible light photoredox catalysis as a means to detect N^2-methylation at guanosine sites by oxidative mutagenesis (Figure 1c).

We first determined the oxidation potential of N^2-methylated guanosines compared with unmodified guanosine using cyclic voltammetry (Figure 2a). The results showed that the oxidation potential of both $m^2{}_2G$ and m^2G were approximately 0.1-0.2V below that of guanosine, providing a potential window for chemoselective oxidation. Using $m^2{}_2G$ as a model, we surveyed various photoredox conditions that were previously shown to be RNA compatible and found that riboflavin under blue light as a photocatalyst and Selectfluor as an oxidant was optimal. Under these conditions, full-conversion of $m^2{}_2G$ occurred within 2 hours, resulting in the formation of the corresponding $m^2{}_2$-Iz, which was confirmed by mass spectrometry (see Figure S1). Formation of $m^2{}_2$-OG or N^2-demethylated byproducts were not observed.

Riboflavin exhibits an excited state reduction potential of E^{T1} = 1.37 V vs. SCE, which could oxidize guanosine in RNA at low levels, potentially causing issues in some sequencing applications. Indeed, guanosine oxidation with riboflavin with higher energy (UVA) photoexcitation has been observed.[22] To test if this undesired process was occurring in our system, we subjected an RNA oligo to the photoredox conditions and analyzed the nucleoside composition by HPLC after nuclease digestion; no guanosine oxidation products were identified (Figure S2). To confirm that guanosine was not undergoing undesired oxidation in more complex RNAs, we performed sequencing to analyze the changes in error at unmodified bases in yeast $tRNA^{Leu}_{CAA}$ after photoredox reaction (Figure 2b). No marked change in error was observed during sequencing analysis in any canonical nucleobase group, supporting the conclusion that significant conversion of guanosine into OG or other mutagenic forms was minimal under the tested conditions. We also examined the stability of RNA under the optimized photoredox conditions and found that a phosphate buffer was essential to minimize RNA degradation, with unbuffered aqueous media resulting in rapid decomposition of the RNA (Figure S3, Table S1). Analysis by RT-qPCR and PAGE showed only 4% loss of RNA compared to the no reaction control (Figure S4).

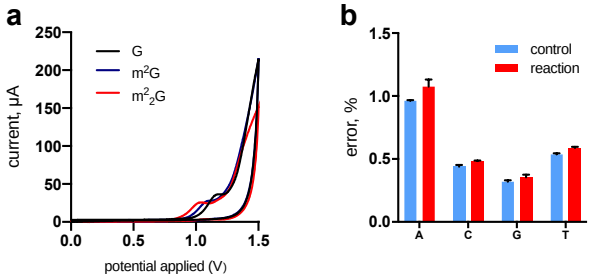

Figure 2. a) Cyclic voltammograms of guanosine (black line), m^2G (blue line) and $m^2{}_2G$ (red line) in 1 M KCl. Scan rate, 10 mV/s, versus Ag/AgCl. b) Comparative error analysis of unmodified nucleobases of $tRNA^{Leu}_{CAA}$ when subjected to photooxidation. Control represents direct sequencing without photoredox catalysis.

Having optimized the photoredox conditions and demonstrated its compatibility with RNA, we sought to examine its chemoselectivity in photooxidative sequencing (PhOxi-seq) of m^2G and $m^2{}_2G$ in various RNA contexts. We first tested PhOxi-seq on yeast $tRNA^{Lys}_{UUU}$, which contains an m^2G and an $m^2{}_2G$ at positions 10 and 26, respectively. tRNA was prepared using the YAMAT-seq[23] workflow to remove the charged amino acids and install sequencing adapters to the no-reaction control and photoredox-treated samples. Following high-throughput sequencing, sequences were trimmed for length and quality, and aligned to the reference sequence using Bowtie 1[24] to enable induced SNP calling.[25] The error at each expected guanosine position was plotted for the control and treated samples (Figure 3a). Consistent with our hypothesis, m^2G_{10} was readily detected with a marked increase in error to 88%. However, surprisingly, $m^2{}_2G_{26}$ experienced a precipitous drop in error from 85% to 16% (Figure 3b). Comparing the error distribution at m^2G_{10} before and after photo-oxidation revealed mostly G→T and G→C transitions, which is consistent with m^2-Iz and m^2-OG formation (Figure 3c). Error analysis at $m^2{}_2G_{26}$ showed conversion from a mostly adenosine incorporation typically observed during read-through of bulky bases,[15] to a more balanced error

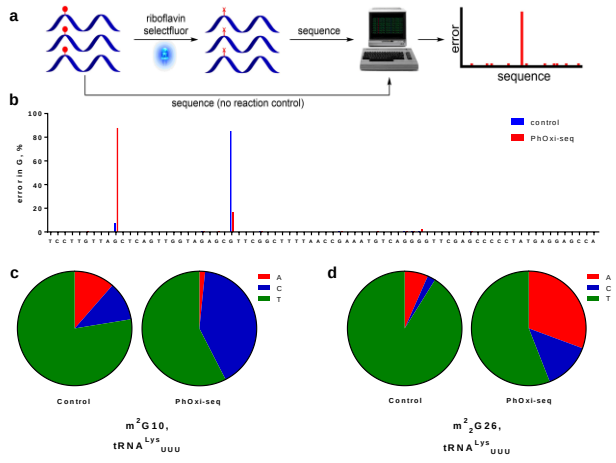

Figure 3. a) Workflow for PhOxi-seq. b) Observed error in each guanosine position in yeast $tRNA^{Leu}_{CAA}$ before (blue) and after (red) photooxidation. Observed change in error distribution at c) m^2G10 and d) $m^2{}_2G26$ during PhOxi-seq.

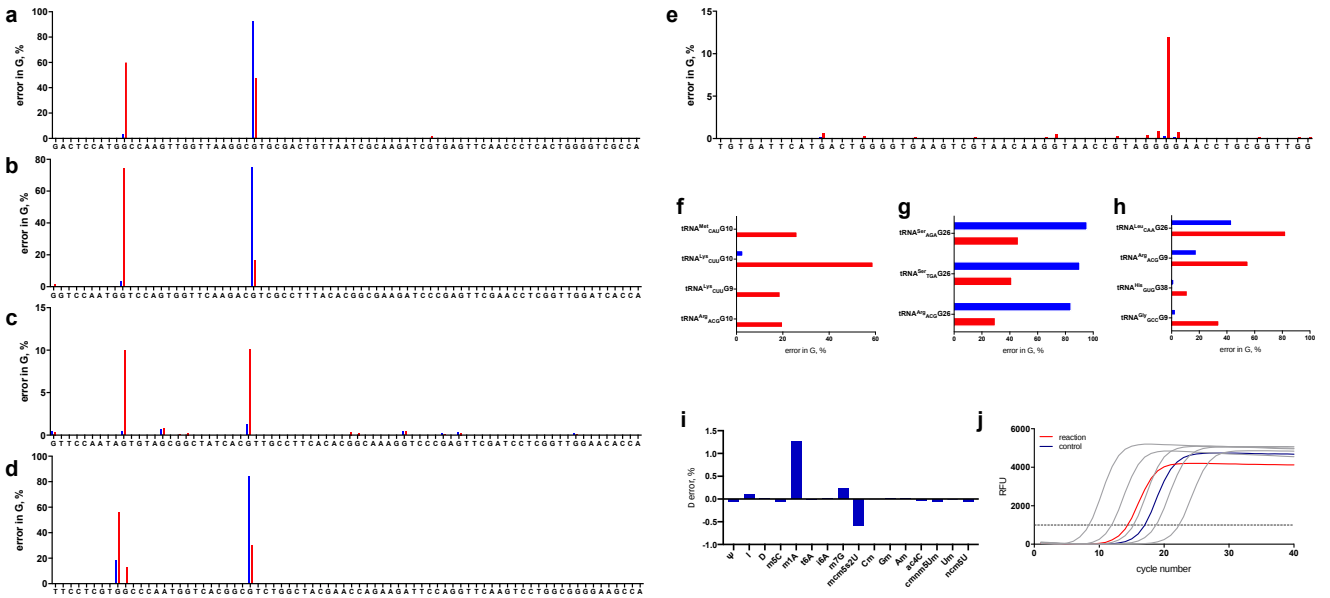

Figure 4. Evaluation of PhOxi-seq on various RNAs. Plots of guanosine error at each position mapped for a) $tRNA^{Asn}_{GUU}$ b) $tRNA^{Val}_{UAC}$ c) $tRNA^{Val}_{CAC}$ d) $tRNA^{Arg}_{ACG}$ e) *E. Coli* 16S rRNA between bases 1471 and 1571; blue bars denote control sequencing run, red bars denote PhOxi-seq run. Guanosine error at known f) m^2G g) $m^2{}_2G$ and h) m^1G sites in yeast tRNA; blue bars denote control sequencing run, red bars denote PhOxi-seq run. i) difference in error between control and PhOxi-seq at various modified bases in tRNA. j) RT-qPCR analysis of read-through efficiency of $m^2{}_2G$ at position 26 in $tRNA^{Phe}_{GAA}$.

distribution (Figure 3d), with no clear base-pairing preference. Encouraged by these results, we examined other readily sequenced yeast tRNAs. $tRNA^{Asn}_{GUU}$ and $tRNA^{Val}_{UAC}$, both of which contain m^2G_{10} and $m^2{}_2G_{26}$ modifications, showed a large increase in error at m^2G sites, with a concomitant decrease in error at $m^2{}_2G$ sites (Figure 4a and b). $tRNA^{Val}_{CAC}$, which contains two instances of m^2G at positions 10 and 26, was effectively sequenced by PhOxi-seq (Figure 4c), and $tRNA^{Arg}_{ACG}$, which contains two consecutive m^2G sites at positions 9 and 10 and an $m^2{}_2G$ located at position 26, was also found to be within the scope of this method (Figure 4d). While there does appear to be a lower response for m^2G_{10} in $tRNA^{Arg}_{ACG}$, PhOxi-seq resulted in a 40-fold increase in error at this position from 0.3% to 13% error. Other known m^2G and $m^2{}_2G$ modifications on yeast tRNAs were also found to be effectively sequenced (Figure 4f,g). We also used PhOxi-seq to detect the m^2G modification at position 1516 of *E. Coli* 16S rRNA with good response and discrimination over flanking guanosines (Figure 4e). While *E. Coli* rRNA was denatured before photooxidation, the observed error was generally lower than those observed in tRNAs, which may be due to the reaction being performed on the whole ribosome. It is likely that larger error responses would be observed for smaller RNA fragments.

PhOxi-seq analysis on other available tRNA modifications showed no major changes in error, suggesting that this method is compatible with a wide scope of functional groups present on RNA (Figure 4i). However, it is important to note that PhOxi-seq alone cannot effectively distinguish between m^2G and m^1G as an increase in error at m^1G sites was observed in all tested cases (figure 4h). While error distributions differ from those seen with m^2G (Figure S5), and native error rates at m^1G sites[26] are significantly higher than those of m^2G, which could enable discrimination between m^1G and m^2G sites, the use of selective m^1G demethylases[26] are likely to be an integral component of PhOxi-seq for N^2-methylation.

During PhOxi-seq analysis, we found that amplification of tRNAs occurred more efficiently after photoredox reaction. It is known that $m^2{}_2G$ is a terminator site during reverse-transcription and can often be challenging to read through.[27] Indeed, enzymatic methods have recently been developed that enable read-through of $m^2{}_2G$ sites during tRNA sequencing.[27] We hypothesized that the conversion of $m^2{}_2G$ into $m^2{}_2$-Iz during PhOxi-seq enables more efficient read-through, perhaps due to the reduced size or *via* alternative base-pairing not available to $m^2{}_2G$. To test this on a model system, we used purified yeast $tRNA^{Phe}_{GAA}$ and monitored read-through of $m^2{}_2G$ by primer extension with and without photooxidation. PAGE analysis revealed that superscript III reverse transcriptase read through $m^2{}_2G$ six times more efficiently after photooxidation (Figure S6), which was confirmed using RT-qPCR (Figure 4j). Such photoredox chemistry could find applications to sequencing of $m^2{}_2G$-containing transcripts as an alternative to engineered polymerases; we are currently exploring the prospect of this approach.

In summary, PhOxi-seq provides a robust approach to the single-nucleotide resolution sequencing of N^2-methylation of guanosine by harnessing the chemoselectivity of photoredox chemistry to enable selective photooxidation of m^2G and $m^2{}_2G$ in the presence of other RNA bases. We have shown that error markedly increases in the case of m^2G by the hypothesized formation of m_2-Iz and m_2-OG that pair with G and T, respectively. Error decreased in the case of $m^2{}_2G$ with concomitant 6-fold enhanced read-through by the formation of $m^2{}_2$-Iz. The method was shown to be compatible within a variety of sequence contexts in both tRNA and rRNA. We anticipate that PhOxi-seq will find

broad use as a straightforward sequencing approach to enable functional studies of N^2-methylation of guanosine in RNA.

ASSOCIATED CONTENT

The Supporting Information, including supporting data, experimental methods, sequencing data, Table S1, and Figures S1-S6 is available free of charge.

AUTHOR INFORMATION

Corresponding Author

* rhili@yorku.ca

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada, Ontario Ministry of Research and Innovation, Canada Foundation for Innovation, and York University.

REFERENCES

(1) Zaccara, S.; Ries, R. J.; Jaffrey, S. R. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 608–624.
(2) Jonkhout, N.; Tran, J.; Smith, M. A.; Schonrock, N.; Mattick, J. S.; Novoa, E. M. *RNA* **2017**, *23*, 1754–1769.
(3) Barbieri, I.; Kouzarides, T. *Nat. Rev. Cancer* **2020**, *20*, 303–322.
(4) Boccaletto, P.; Machnicka, M. A.; Purta, E.; Piątkowski, P.; Bagiński, B.; Wirecki, T. K.; de Crécy-Lagard, V.; Ross, R.; Limbach, P. A.; Kotter, A.; Helm, M. *Nucl. Acids Res.* **2017**, *46*, D303–D307.
(5) Li, X.; Xiong, X.; Yi, C. *Nat. Meth.* **2017**, *14*, 23-31.
(6) Jurga, S.; Barciszewski, J. *Epitranscriptomics*; Springer International Publishing (Verlag), 2021
(7) Sergiev, P. V.; Bogdanov, A. A.; Dontsova, O. A. *Nucl. Acids Res.* **2007**, *35*, 2295–2301.
(8) Chen et al. *Science* **2016**, *351,* 6271.
(9) Zhang et al, *Nat. Cell Biol.* **2018**, *20*, 535–540.
(10) Blaesius, K. et al. *Am. J. Med. Genet. A* **2018**, *176*, 2517–2521.
(11) Najmabadi et al. *Nature* **2011**, *478*, 57–63.
(12) Dai, Q.; Zheng, G.; Schwartz,M. H.; Clark, W. C.; Pan, T. *Angew. Chem. Int. Ed.* **2017**, *56*, 5017–5020.
(13) Chen, W.; Song, X.; Lv, H.; Lin, H. *Mol. Ther. Nucl. Acid* **2019**, *18*, 253–258.
(14) Ao, C.; Zou, Q.; Yu, L. *Methods*, in press, **2021**, https://doi.org/10.1016/j.ymeth.2021.05.016.
(15) Ryvkin, P.; Leung, Y. Y.; Silverman, I. M.; Childress, M.; Valladares, O.; Dragomir, I.; Gregory, B. D.; Wang, L-S. *RNA* **2017**, *19*, 1684–1692.
(16) Kobayashi, K.; Tagawa, S. *J. Am. Chem. Soc.* **2003**, *125*, 10213–10218.
(17) Morikawa, M.; Kino, K.; Oyoshi, T.; Suzuki, M.; Kobayashi, T.; Miyazawa, H. *RSC Avd.* **2013**, *3*, 25694–25697.
(18) Kino, K.; Sugiyama, H. *Chem. Biol.* **2001**, *8*, 369–378.
(19) Xie, L.-I.; Yang, Y.-T.; Wang, R.-L.; Cheng, H.-P.; Li, Z.-Y.; Liu, L.; Mao, L.; Wang, M.; Cheng, L. *Angew. Chemie. Int. Ed.* **2019**, *58*, 5028–5032.
(20) Nappi, M.; Hofer, A.; Balasubramanian, S.; Gaunt, M. J. *J. Am. Chem. Soc.* **2020**, *142*, 21484–21492.
(21) Patel, S.; Badir, S. O.; Molander, G. A. *Trends Chem.* **2021**, *3*, 161–175.
(22) Kino, K.; Kobayashi, T.; Arima, E.; Komori, R.; Kobayashi, T.; Miyazawa, H. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 2070-2074.
(23) Shigematsu, M.; Honda, S.; Loher, P.; Telonis, A. G.; Rigoutsos, I.; Kirino, Y. *Nucl. Acids Res.* **2017**, *45*, e70.
(24) Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S. L. *Genome Biol.* **2009**, *10*, R25.
(25) https://github.com/ckorfmann/snp-counter
(26) Wang, Y.; Katanski, C. D.; Watkins, C.; Pan, J. N. Dai, Q.; Jiang, Z.; Pan, T. *Nucl. Acids Res.* **2021**, *49*, e30.
(27) Dai, Q.; Zheng, G.; Schwartz, M. H.; Clark, W. C.; Pan, T. *Angew. Chemie. Int. Ed.* **2017**, *56*, 5017–5020.