Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution

Vahid Khoddami[a,b,c,1,2], Archana Yerra[b,c,1], Timothy L. Mosbruger[d], Aaron M. Fleming[e], Cynthia J. Burrows[e,3], and Bradley R. Cairns[b,c,3]

[a]Department of Cell Biology, Harvard Medical School, Boston, MA 02115; [b]Howard Hughes Medical Institute, University of Utah School of Medicine, Salt Lake City, UT 84112; [c]Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT 84112; [d]Bioinformatics Shared Resource, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT 84112; and [e]Department of Chemistry, University of Utah, Salt Lake City, UT 84112

Contributed by Cynthia J. Burrows, January 25, 2019 (sent for review October 9, 2018; reviewed by Juan D. Alfonzo, Thomas Carell, and Peter C. Dedon)

The breadth and importance of RNA modifications are growing rapidly as modified ribonucleotides can impact the sequence, structure, function, stability, and fate of RNAs and their interactions with other molecules. Therefore, knowing cellular RNA modifications at single-base resolution could provide important information regarding cell status and fate. A current major limitation is the lack of methods that allow the reproducible profiling of multiple modifications simultaneously, transcriptome-wide and at single-base resolution. Here we developed RBS-Seq, a modification of RNA bisulfite sequencing that enables the sensitive and simultaneous detection of m^5C, Ψ, and m^1A at single-base resolution transcriptome-wide. With RBS-Seq, m^5C and m^1A are accurately detected based on known signature base mismatches and are detected here simultaneously along with Ψ sites that show a 1–2 base deletion. Structural analyses revealed the mechanism underlying the deletion signature, which involves Ψ-monobisulfite adduction, heat-induced ribose ring opening, and Mg^{2+}-assisted reorientation, causing base-skipping during cDNA synthesis. Detection of each of these modifications through a unique chemistry allows high-precision mapping of all three modifications within the same RNA molecule, enabling covariation studies. Application of RBS-Seq on HeLa RNA revealed almost all known m^5C, m^1A, and ψ sites in tRNAs and rRNAs and provided hundreds of new m^5C and Ψ sites in noncoding RNAs and mRNAs. However, our results diverge greatly from earlier work, suggesting ∼10-fold fewer m^5C sites in noncoding and coding RNAs and the absence of substantial m^1A in mRNAs. Taken together, the approaches and refined datasets in this work will greatly enable future epitranscriptome studies.

RNA modification | pseudouridine | RNA methylation | m^1A | methyl adenosine

Covalent modifications of RNA are numerous (1), and transcriptome-wide profiling enables broad and systematic analyses (2–4). Thus far, transcriptome-wide profiling has been reported for a limited number of modifications including N^6-methyladenosine (m^6A), 5-methylcytosine (m^5C), pseudouridine (Ψ), and N^1-methyladenosine (m^1A) (5–14). However, profiling methods that provide sensitive and true single-base resolution are currently available only for m^5C (9, 14, 15) and m^1A (16); three of these (m^6A, m^1A, and Ψ) have involved initial enrichment or detection via antibodies (for m^6A or m^1A) (5, 6, 8, 10) or by techniques involving polymerase pausing/termination during reverse transcription (for m^1A and Ψ) (7, 11–13, 17, 18). Recent single-base techniques for Ψ (19) rely on a bulky adduct formation before detection. Furthermore, although the current methods for Ψ profiling are useful, most lack the sensitivity, resolution, and technical ease needed for widespread adoption or straightforward candidate site validation (7, 11–13, 20). To provide simultaneous detection of m^5C, m^1A, and Ψ at single-base resolution transcriptome-wide from the same sample, we developed a molecular approach and analysis pipelines for Ψ and improved sequencing-based methods for m^5C and m^1A. First, we provide the conceptual basis for sequencing/mismatch-based detection of all three modifications (Fig. 1*A*) and an example tRNA (glycine) that illustrates modification clarity within our HeLa cell dataset (Fig. 1 *B* and *C*, with multiple additional examples in *SI Appendix*, Figs. S1–S3).

Detection of m^5C in RNA (and DNA) relies on differential sensitivity to bisulfite: unmethylated cytosine is efficiently deaminated by bisulfite ions converting cytosine to uridine, which is subsequently read as thymidine following desulfonation, RT-PCR, and sequencing. In contrast, m^5C resists bisulfite and remains cytosine after sequencing (15, 21) (Fig. 1*A*). We improved prior m^5C profiling methods by combining heat and the strong chemical denaturant formamide, which improves RNA denaturation and bisulfite treatment (which preferentially modifies single-stranded RNA), providing a global C → T conversion

Significance

The field of RNA modification would be significantly advanced by the development of sensitive, accurate, single-base resolution methods for profiling multiple common RNA modifications in the same RNA molecule. Our work provides several advances toward that goal, including (*i*) quantitative methods for profiling Ψ sites at true base-pair resolution transcriptome-wide, (*ii*) a chemical understanding of our observed Ψ-dependent deletion signature, (*iii*) improved methods for profiling m^5C and m^1A, and (*iv*) a coupling of these methods for the simultaneous detection of all three modifications in the same RNA. Together, the combinatorial ability and relative ease of execution provided by this procedure should greatly forward epitranscriptome studies involving these three very common RNA modifications.

Author contributions: V.K. and B.R.C. designed research; V.K., A.Y., and A.M.F. performed research; V.K., A.Y., T.L.M., and A.M.F. contributed new reagents/analytic tools; V.K., A.Y., T.L.M., A.M.F., C.J.B., and B.R.C. analyzed data; and V.K., C.J.B., and B.R.C. wrote the paper.

Reviewers: J.D.A., Ohio State University; T.C., Ludwig Maximilians University; and P.C.D., Massachusetts Institute of Technology.

The authors declare no conflict of interest.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, https://www.ncbi.nlm.nih.gov/geo (accession no. GSE90963). All custom computer scripts reported in this paper have been deposited in GitHub, https://github.com/HuntsmanCancerInstitute/RBSSeqTools.

[1]V.K. and A.Y. contributed equally to this work.

[2]Present address: Department of Stem Cells and Developmental Biology, Cell Science Research Center, Royan Institute for Stem Cell Biology and Technology, Academic Center for Education, Culture and Research, 16635-148 Tehran, Iran.

[3]To whom correspondence may be addressed. Email: burrows@chem.utah.edu or brad.cairns@hci.utah.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1817334116/-/DCSupplemental.

Published online March 14, 2019.