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.

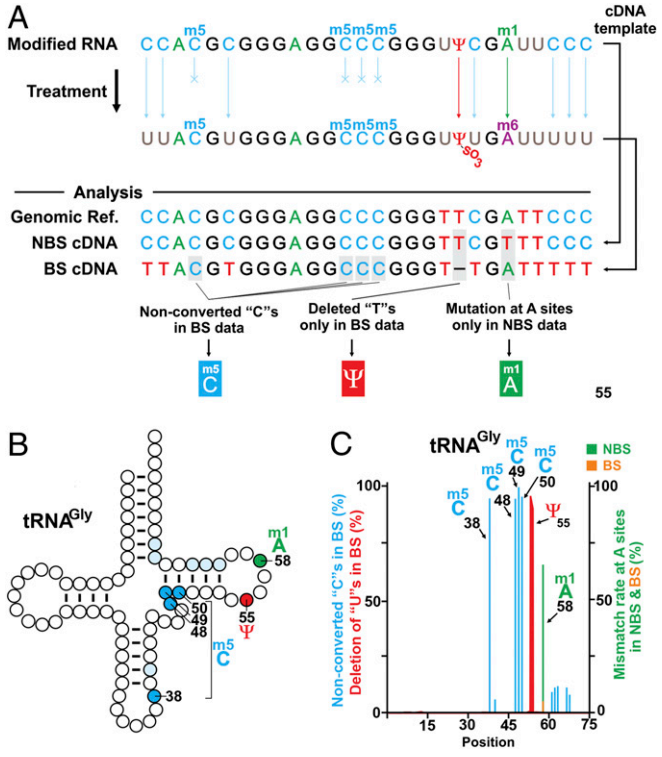

Fig. 1. RBS-Seq enables simultaneous base-pair resolution transcriptome-wide mapping of m^5C, m^1A, and Ψ. (*A*) Schematic of reactivity of modified nucleotides to bisulfite (*Top*) and the principle of simultaneous mapping of modified nucleotides (*Bottom*). For m^5C, bisulfite treatment deaminates Cs (converting to Us; Ts upon cDNA sequencing), whereas m^5Cs resist bisulfite (remain Cs upon cDNA sequencing) establishing sites of cytosine methylation. For m^1A, cDNA synthesis at sites with m^1A often confers misincorporation/mismatch in the NBS sample. In contrast, during treatment, m^1A becomes converted to m^6A via Dimroth rearrangement (methyl passes from N^1 to N^6), which faithfully templates thymine (cDNA remains adenosine) in the BS sample. Thus, comparison of NBS and BS samples identifies m^1A sites. For Ψ, Ψ nucleotides upon bisulfite treatment form a stable monobisulfite adduct (Fig. 4) causing frequent bypass with reverse transcriptase, leaving a deletion signature at the exact modified sites, evident exclusively in the BS samples. (*B*) Schematic representation of $tRNA^{Gly}$, indicating its well-known multiple m^5C and single m^1A and Ψ modified sites. (*C*) Bar graph summarizing the actual RBS-Seq results from HeLa cell line for a $tRNA^{Gly}$ locus indicating the exact locations of the modified nucleotides and their levels. The low levels of m^5C shown at positions 40 and 60–66 have been shown previously for a subset of tRNA types (31, 32).

frequency of 99.7% in HeLa RNA (*SI Appendix*, Figs. S4–S8). Optimization was quantified via synthetic RNA oligomers, with m^5C bases placed within and/or outside of regions of secondary structure (*SI Appendix*, Fig. S5). We applied RBS-Seq to HeLa RNA, using three types of input RNA species [polyA-selected (~85 M reads), rRNA-depleted (~200 M reads), and small RNA (~92 M reads)] (22), and our analysis pipelines (*SI Appendix*, Fig. S9) (23) compared datasets derived from bisulfite-treated (BS) and non–bisulfite-treated (NBS; untreated) RNAs of the same sample, a regimen which reduces false positives generated by incorrect alignments resulting from reduced nucleotide complexity. We then aggregated results from all three input types and filtered out additional false positives via computational and visual inspection for C/G tracts and strong secondary structure and imposed thresholds for nonconversion [≥20% (FDR ≤ 0.05)] (*SI Appendix*, Fig. S9). This combination of approaches and thresholds yielded a list of high-threshold candidate sites in the following RNA categories: 486 total m^5C sites, representing 297 unique sites in abundant noncoding RNAs (tRNAs and rRNA, together), 143 sites in mRNAs (e.g., PTEN and HDGF; Figs. 2 *A* and *B* and 3*A*, *SI Appendix*, Fig. S2*A*, and Datasets S1 and S2), 14 pseudogenes, and 32 other noncoding RNAs. New sites within prominent mRNAs include PTEN, XRCC3, FANCA, RXRB, FGFR4, and EIF3B (Dataset S1). Importantly, examination of known/validated sites in tRNAs demonstrated that RBS-Seq has a dynamic range for m^5C approaching 100% at single cytosines (e.g., C49 in Fig. 1*C*). Although our read numbers exceeded prior studies, our yield of 486 high-threshold candidate m^5C sites in mRNA was far lower than the 10,275 sites reported previously (14), largely due to our more effective denaturation and deamination/conversion, lowering false positives (*SI Appendix*, Figs. S10 and S11, and Dataset S1). In keeping, more recent m^5C profiling in mouse ESCs that applies additional statistical and analytical parameters to remove false-positives reports 266 sites in mRNA passing thresholds (24).

Unlike m^6A, m^1A compromises A:T Watson–Crick base pairing, which pauses reverse transcriptase and elicits frequent nucleotide misincorporation, generating a single-base mismatch signature useful for m^1A identification (8, 17, 25). As expected, in our NBS datasets from RBS-Seq, we indeed detected significant (FDR < 0.01) m^1A-related mismatches at well-known m^1A sites in noncoding RNAs [e.g., m^1A-1322 in 28S rRNA (Fig. 2*C* and *SI Appendix*, Fig. S2*B*) and m^1A-58 in all tRNAs (Fig. 2*D* and Dataset S3)]. Unexpectedly, these mismatches were wholly absent or greatly diminished in our BS datasets (*SI Appendix*, Table S3), with tRNAs displaying the remarkable dynamic range of our method (~90% for $tRNA^{Gly}$ and $tRNA^{Thr}$; Fig. 2*D*). Regarding the basis, conversion of m^1A to m^6A (involving transfer of the methyl group from N^1 to the N^6 position) occurs through a well-studied process known as the Dimroth rearrangement (26) (*SI Appendix*, Fig. S12), which readily occurs in the alkaline heat conditions present in the desulfonation step of the RBS-Seq procedure (*SI Appendix*, Fig. S7). Thus, comparisons of base mismatch frequency within a BS sample compared with its matched original NBS sample reveals sites of m^1A (*SI Appendix*, Fig. S13 and Dataset S3). Notably, by our methods and analyses, no significant m^1A (>1% at individual A sites) was detected at any single site within an mRNA (*SI Appendix*, Figs. S14 and S15, and Dataset S4). Our results contrast with initial studies claiming thousands of m^1A sites in mRNAs (10, 25) but are corroborated by more recent studies, which quantify m^1A in mRNAs and lncRNAs as extremely rare (15 total sites in HEK293T cells) (24, 27–29).

We then focused considerable attention on Ψ. Fortuitously, we observed a reproducible highly penetrant 1–2 nucleotide deletion signature at virtually all known Ψ sites in tRNAs, rRNAs, and snRNAs, exclusively in BS samples. Notably, because our approach does not stop reverse transcriptase, it can uniquely reveal two nearby Ψ sites on the same RNA (*SI Appendix*, Figs. S2*C*, S16–S19 and Dataset S5). Regarding penetrance, 47 uniquely mapping, known/validated Ψ sites in tRNAs (from prior studies, including Ψ55) displayed >50% penetrance (e.g., >90% for $tRNA^{Gly}$; Fig. 1 *B* and *C*), demonstrating the exceptional dynamic range of RBS-Seq for Ψ detection (Dataset S5). Below we address in more detail the deletion mechanism; however, this consistent and unique feature motivated expansion to identify novel Ψ sites transcriptome-wide, and for comparative purposes we chose HeLa cells. Here a custom analysis pipeline was developed involving a statistical approach (*SI Appendix*, Fig. S16), which revealed 754 unique sites: 388 sites in various noncoding RNA species, 322 sites in mRNAs, and 44 sites in pseudogenes (including CDC6; Fig. 2 *E* and *F*, *SI Appendix*, Fig. S2*C*, and Dataset S5; FDR < 0.001) with a strong bias for coding regions (Fig. 3*B*). Thus, our work provides hundreds of Ψ sites in noncoding and mRNA species, which show clear enrichment for GO categories related to protein translation and metabolism (especially RNA metabolism) (*SI Appendix*, Fig. S17, and Dataset S6).

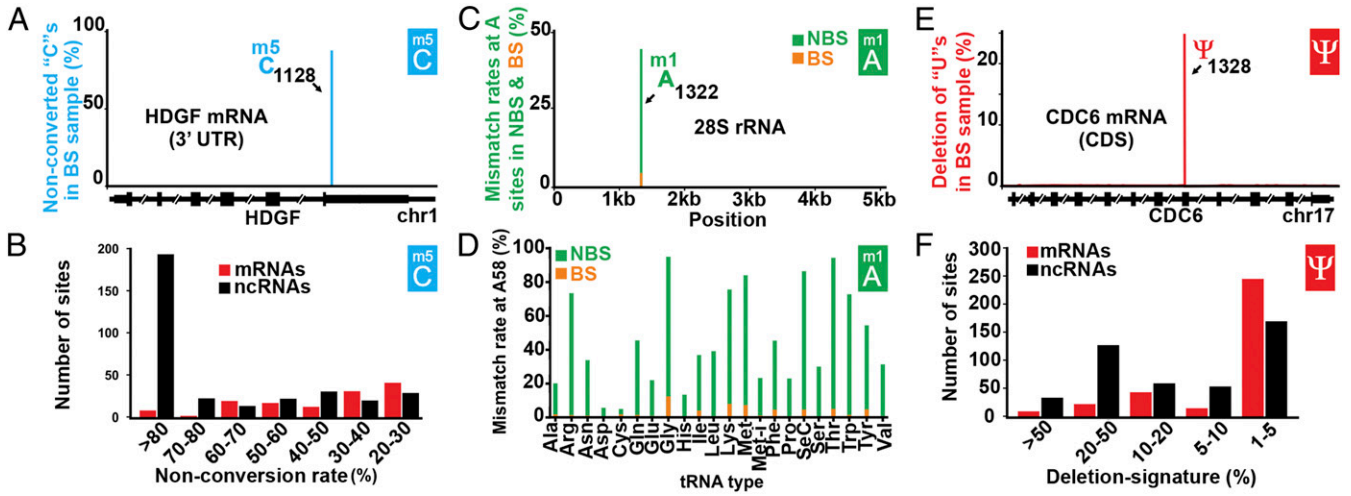

Fig. 2. Transcriptome-wide RBS-Seq analysis in HeLa cells shows widespread m^5C and Ψ in mRNAs and ncRNAs and m^1A almost exclusively in rRNAs and tRNAs. (*A*) HDGF mRNA bears a single m^5C base, within its 3′ UTR. (*B*) Distribution and dynamic range of m^5C sites in coding and noncoding RNAs, stratified by the nonconversion rate (percent of reads that contain a single cytosine that remains cytosine after bisulfite treatment). (*C*) Mismatch rate (percent of reads with a base mismatch) at human 28S rRNA transcript reveals a single well-known m^1A nucleotide at position 1322. (*D*) Mismatch rate at the conserved well-known A58 sites of human tRNAs (from HeLa cells) in NBS and BS samples. Representatives of each tRNA type with the highest mismatch rate are shown. (*E*) Deletion rate (percent of reads bearing a 1–2 base deletion) along CDC6 mRNA reveals a single Ψ nucleotide. (*F*) Distribution and dynamic range for Ψ candidate sites in coding and noncoding RNAs.

Sites in prominent mRNAs include SMC2, EIF3D, POLE4, LMO7, CCDC22, ATP5F1, and TRIM8 (Dataset S5).

Four groups have previously conducted transcriptome-wide Ψ profiling reports under different names: Pseudoseq (7) and Ψ-seq (13) (using yeast and mammalian cells), PSI-seq (12) (yeast), and CeU-Seq (11) (mammalian cells). All four methods share the same principle: treatment of RNA with the chemical *N*-cyclohexyl-*N*′-(2-morpholinoethyl)-carbodiimide-metho-*p*-toluenesulfonate (CMC), which leaves a bulky group on pseudouridines, stopping reverse transcriptase during cDNA synthesis, thus indicating sites of pseudouridylation globally via RNASeq. These CMC-based techniques have proven useful for identifying Ψ sites, especially in high-abundance RNAs, and for identifying candidate Ψ sites in mRNAs. However, despite utilizing similar methods and identical yeast strains, overlap between candidate sites was extremely low (typically <4%) and was equivalently low when we compared published CMC-based results in mammalian cell types (typically HeLa and/or HEK293) (20) (*SI Appendix*, Fig. S20, and Dataset S7). Interestingly, candidate Ψ sites from RBS-Seq using HeLa cells overlapped better with prior CMC-based studies (either HeLa or HEK293) than did the prior studies with themselves (*SI Appendix*, Fig. S21, and Dataset S7), consistent with RBS-Seq revealing a higher proportion of positives.

To better understand these differences, we turned to validation approaches. CMC-based methods that rely on cDNA chain termination for mapping Ψ sites present challenges for validation, requiring quantities of pure target RNA beyond feasibility for most mRNAs; thus, most prior studies either lacked validation (7, 13) or verified only very few (fewer than five) highly abundant candidate sites (11, 12). RBS-Seq, in contrast, provides a straightforward high-throughput validation protocol easily adapted to mRNA because the Ψ-dependent deletion signatures that appear within the corresponding cDNAs are easily scaled up through gene-specific PCR amplification and quantified by high-throughput sequencing of barcoded amplicons. Thus, for validation and comparisons to prior studies, we tested 60 candidate sites, which we partitioned into four groups: group I, sites uniquely identified by RBS-Seq (12 sites); group II, sites shared between RBS-Seq and at least one CMC-based method (25 sites); group III, sites detected in at least one of the CMC-based methods but not identified by RBS-Seq due to falling below our read coverage thresholds (14 sites); and group IV, sites from one or more CMC-based methods but not RBS-Seq, despite sufficient coverage in RBS-Seq (nine sites) (Dataset S8). For validation tests, we treated HeLa or HEK293 total RNA to a streamlined bisulfite + heat + $MgCl_2$ protocol via chemistry described below followed by RT-PCR involving barcoded ∼125-bp amplicons (on average), sequencing (termed RBS-MiSeq), and our deletion signature analysis pipeline (*SI Appendix*, Fig. S22*A*). Notably, the vast majority (34 of 37) of group I and II sites validated, yielding a clear deletion signature involving tens of thousands of sequenced reads with HeLa and/or HEK293 datasets, providing confidence in sites identified by RBS-seq. For group III, 10 of 14 validated, suggesting that increasing the sequencing depth and/or applying our focused validation approach can resolve the rare false negatives generated by RBS-Seq. Finally, none of the candidates in group IV (0 of 9) validated, strongly suggesting a much higher false positive rate with any of the CMC-based techniques alone compared with RBS-Seq, consistent with their low overlap in prior studies. Examples of each group are provided in *SI Appendix*, Fig. S22*B*, and complete results are provided in Dataset S8.

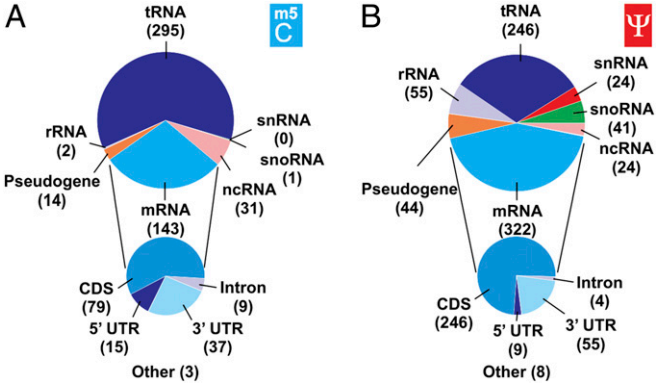

Fig. 3. (*A*) Distribution of candidate m^5C sites in RNA species, with expanded mRNA annotations. (*B*) Distribution of candidate Ψ sites in different RNA species, with expanded mRNA annotations. See Datasets S1 and S5 for all sites and their corresponding annotations.

To independently test our Ψ profiling approaches and results, we examined Dyskerin (DKC1), the most disease-relevant human Ψ synthase. Mutation of *DKC1* causes dyskeratosis congenita, characterized by short telomeres and bone marrow failure (30, 31). DKC1 utilizes H/ACA box snoRNAs to guide Ψ targeting to rRNAs via base pairing between the snoRNA and the target rRNA (32), and DKC1 also interacts with telomerase noncoding RNA (TERC) (33, 34), but whether TERC receives substantial Ψ is uncertain (35). To resolve this issue, high-throughput RBS-Seq followed by deletion signature analysis was performed on both total and polyA-selected RNAs isolated from DKC1-depleted HeLa cell via siRNAs, yielding ~84% reduction in DKC1 transcript levels (*SI Appendix*, Fig. S22 *C* and *D*). Comparison of the DKC1-siRNA with control-siRNA datasets revealed significant reduction (>25% reduction, FDR < 0.01) of the deletion signature levels in 227 sites; most reside within rRNAs, although 18 sites were observed within mRNAs (*SI Appendix*, Fig. S22 *E* and *F*, and Dataset S9). Curiously, the 58 DKC1-dependent sites in HEK293 mRNAs reported by Ψ-seq show no overlap with the 18 sites found in HeLa mRNAs by RBS-Seq. Moreover, because Ψ-seq but not RBS-Seq reported two DKC1-dependent Ψ sites within TERC in HEK293 cells (13), we specifically tested TERC at both sites with our streamlined RBS-MiSeq validation procedure in both HeLa and HEK293 cells. Notably, despite over 30 K reads overlapping both sites in both cell types, no significant deletion was observed (Dataset S8), suggesting that TERC is an interacting partner of DKC1 but not a direct pseudouridylation substrate in these cell types under the conditions tested.

Finally, to elucidate the chemistry of the observed 1–2 base deletion signature at Ψ and to guide validation methodologies, we determined which step(s) of our RBS-Seq protocol elicited base deletion by utilizing a synthetic 70-mer oligonucleotide bearing two Ψ sites and quantifying base deletion frequencies. Strikingly, bisulfite treatment alone failed to induce any deletion signature, whereas heating (75 °C) the BS RNA in the presence of magnesium ions (20 mM) for 15 min was both necessary and sufficient for generating the penetrant deletion signature (Fig. 4 *A*–*C*

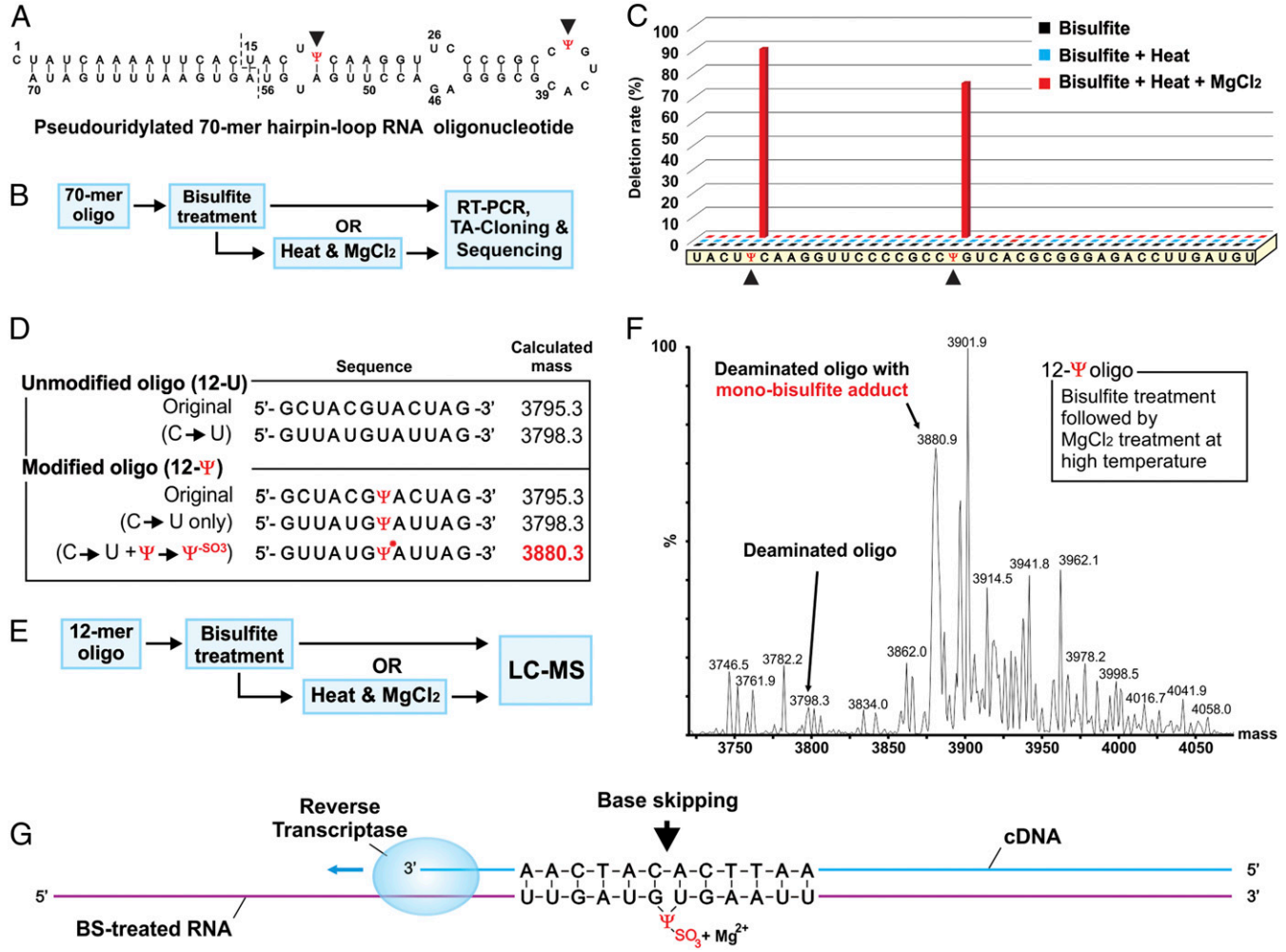

Fig. 4. Characterization of a pseudouridine-bisulfite adduct and heat/Mg^{2+}-induced rearrangement to elicit reverse transcriptase bypass. (*A*) Sequence and intramolecular folding of pseudouridylated 70-mer RNA oligonucleotide used in the downstream experiments. The two Ψ sites (in red) are indicated by arrowheads. (*B*) Flowchart of oligo treatments, RT-PCR, TA cloning, and Sanger sequencing of individual colonies. (*C*) Summary of the deletion signatures obtained from oligonucleotide experiments with the reference sequence and the two Ψ sites at the bottom, showing the insufficiency of the bisulfite step, and requirement for the subsequent heat + $MgCl_2$ step to generate the deletion signatures. (*D*) Sequence and calculated mass for 12-mer control (12-U) and pseudouridylated (12-Ψ) oligomers used in the downstream experiments. (*E*) Reaction sequence and methods used for Ψ reactivity studies with 12-mers. (*F*) Mass spectrum for 12-Ψ after bisulfite and subsequent heat + $MgCl_2$ treatments shows formation of a stable monobisulfite adduct. Mass spectrum for the 12-U and 12-Ψ with only bisulfite treatment is provided in *SI Appendix*, Fig. S24. (*G*) A proposed model showing that during cDNA synthesis, ribose ring-opened Ψ-monobisulfite is oriented away from the polymerization site, reinforced by Mg^{2+}, explaining base skipping.

and *SI Appendix*, Figs. S24–S27). Similar treatments with a synthetic 12-mer RNA oligonucleotide containing a single Ψ followed by MS analysis displayed a major peak consistent with a stable, monoadduct of bisulfite (Fig. 4 *D–F* and *SI Appendix*, Fig. S28). Next, the site of covalent attachment of the bisulfite group was determined by exposing the Ψ nucleoside itself to the reaction sequence followed by structural analysis (*SI Appendix*, Fig. S23*A*). Notably, the nucleoside reaction analyzed by LC-MS revealed two stable monobisulfite adducts (*SI Appendix*, Figs. S29 and S30). The UV-vis for these compounds displayed $\lambda_{max} \sim$ 260 nm (*SI Appendix*, Fig. S31), supporting the aromaticity of the base remaining intact after the reaction. On the basis of ^{1}H-NMR, the bisulfite attachment was determined to be on the 1′ carbon, which involved an opening of the ribose ring to yield a pair of isomeric products, consistent with prior preliminary results (*SI Appendix*, Fig. S32) (36). We propose initial addition of bisulfite to the base, followed by a heat-induced migration of the bisulfite to the ribose, yielding rearomatization and the formation of a stable ring-opened sugar adduct (*SI Appendix*, Fig. S23*B*). Notably, previous studies examining polymerase bypass of similar ring-opened template sites reported a strong tendency for deletions (37), consistent with our observations and interpretation that the ribose ring-opened bisulfite adduct of pseudouridine underlies the deletion signature. Interestingly, although Mg^{2+} was an absolute requirement for generation of the deletion signature assessed by cDNA sequencing (Fig. 4*C*), it proved dispensable for generating the ring-opened sugar adduct in the nucleoside experiments, suggesting that Mg^{2+} helps reorient the ribose ring-opened Ψ bisulfite adduct away from the polymerization site or stabilizes a preexisting configuration that eventually causes polymerase bypass (Fig. 4*G*).

Discussion

This work provides five advances in RNA modification profiling: (*i*) improved methods for profiling m^5C and m^1A; (*ii*) quantitative methods for profiling Ψ sites at true base resolution transcriptome-wide; (*iii*) a chemical understanding of the Ψ-dependent deletion signature; (*iv*) a coupling of these methods for the simultaneous detection of all three modifications in the same sample, which has provided hundreds of candidate sites of modification; and (*v*) a streamlined Ψ candidate site validation procedure for bulk verification of dozens of candidate sites in the same sample. 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, and the refined lists of high-threshold mapped sites in HeLa cells should enable better-focused downstream functional studies. Furthermore, because RBS-Seq also provides transcript abundance (like a typical RNAseq), the combined outputs present a multidimensional (4D) dataset that may prove useful both for basic investigations and diagnostic settings.

Methods

Detailed methods are provided in *SI Appendix, SI Methods*.

Cell Culture (Including siRNA Treatment). HeLa cells were seeded in 100-mm plates at 2×10^6 per plate in DMEM (Gibco) containing 4.5 g/L D-glucose, L-glutamine, and 110 mg/L sodium pyruvate and supplemented with 10% FBS. At ∼75% confluency, cells were harvested via TrypLE Express (Gibco) and washed once with 1× PBS, pH 7.4 (Gibco).

For the DKC1 depletion experiment, siRNA treatment was performed in two sequential transfections per sample, 72 h apart. HeLa cells were seeded at 3×10^5 per well and transfected with Lipofectamine RNAiMAX (Invitrogen) and 60 pmol of siRNA (either Dharmacon's siGENOME Human DKC1 siRNA for the DKC1 knockdown or Dharmacon's siGENOME nontargeting siRNA pool no. 1 for the control sample). Cells were split and reseeded 48 h after the first transfection and harvested 72 h after the second transfection.

RNA Isolation and Preparation. Total RNA from the above samples was isolated using TRIzol Reagent (Invitrogen) and split. rRNA depleted samples were obtained via RiboMinus Transcriptome Isolation Kit for human/mouse (Ambion). The polyA-selected sample was isolated from total RNA using polyA Spin mRNA Isolation Kit (New England Biolabs). Small RNA (enriching transcripts <200 bp) was isolated using mirVana miRNA Isolation. See *SI Appendix, SI Methods*, for RNA fragmentation.

Bisulfite Treatment. Processed RNA was denatured by incubating 45 μL RNA (5 μg) in 240 μL deionized formamide at 95 °C for 5 min before chilling on ice. For the sulfonation step, 312 μL of freshly prepared 5 M sodium bisulfite (pH 5) with 3 μL 100 mM hydroquinone was added to each denatured sample and incubated at 50 °C. To desulfonate, each sample was purified on Illustra NAP-10 columns (GE) and incubated in 2 M Tris buffer, pH 9.0 0 (Trizma Preset crystals, pH 9.0; Sigma-Aldrich), for 2 h at 37 °C. The RNA was recovered by ethanol precipitation.

Library Preparation and Sequencing. For the transcriptome-wide study and for the DKC1 depletion experiment, we used the Illumina Tru-Seq Small RNA kit to generate paired-end libraries. The resulting libraries were sequenced in a 101-cycle paired-end format on an Illumina HiSeq 2000 instrument.

Validation of Candidate Ψ Sites by RBS-MiSeq. PCR amplification yielded ∼300 bp regions surrounding each of 60 candidate sites via a primer design compatible with bisulphite, in which all unmethylated Cs have been converted to Ts. In addition, two DCK1-dependent sites in TERC RNA were tested. HeLa and HEK293 cells were cultured as described above, and RNA was extracted and bisulfite-treated, incubated sequentially in library preparation buffers lacking enzyme, and recovered by ethanol precipitation. Random hexamers and SuperScript II Reverse Transcriptase (Invitrogen) generated cDNA, from which we generated PCR amplicons using our PCR primer sets, which were then pooled and used for MiSeq DNA library preparation using different barcodes for HeLa or HEK293 sets according the Illumina protocols. Libraries were sequenced on a MiSeq instrument.

Bioinformatics Methods. Transcriptome-wide sequencing reads from BS, NBS, and the DKC-1 experiments were aligned using Novoalign (Novocraft) to standard and bisulfite reference index of hg19 chromosome, scaffold and splice junction sequences, accommodating repeat reads, and trimming adaptor sequences. Reads mapping to certain small and repetitive RNA classes (tRNA, rRNA, snRNA, and snoRNA) were extracted and realigned to a custom reference containing only unique representative sequences of the above. All alignment files were processed identically. The processed alignments were then analyzed using custom scripts (https://github.com/HuntsmanCancerInstitute/RBSSeqTools) to generate tables of candidate sites for each individual modification based on the criteria as listed. For m^5C, we selected only those sites from all reference "C" positions which had a read depth ≥10 in both BS and NBS datasets, a C->T nonconversion rate of ≥20% in the BS dataset, and an FDR ≤ 0.05 for nonconversion; sites were further screened manually for evidence of nonconversion arising from undenatured structure or mapping errors. For m^1A, all reference A positions having coverage of at least 100 reads, positions were selected where the non-A rate of the NBS sample was significantly higher than the BS sample; two of C, G, or T were significantly higher than the error rate; and the FDR of C was not lower than that of both G and T. Additionally, to compare our data to publicly available m^1A datasets, we screened our data for previously reported positions and filtered using the same criteria for m^1A. Because previously available data were reported as enriched regions, these were merged across replicates, and the RefSeq transcript positions were converted to genomic positions using the python hgvs package. The entire region was scanned within our data for potential m^1A sites. Pseudouridine positions were called if the following criteria were met: bisulfite (BS) ≥5 deletions, BS fraction deletion ≥0.01, and BS coverage ≥10 reads. Adjacent positions were merged into individual deletion groups, which were further pruned by removing positions with fraction deletions less than half the maximum observed faction deletion in the group. See *SI Appendix, Methods*, for further details on pruning.

Investigating the Source of the Deletion Signature for Ψ. A synthetic 12-mer RNA strand (10 nmols), 5′-GCU ACG ΨAC UAG-3′, was bisulfite treated (as described above) and dialyzed against ddH_2O for 36 h at 4 °C, and the water was changed every 8 h. After the 36-h dialysis, the sample was then dialyzed against ddH_2O containing 3 mM NH_4OAc for 36 h at 4 °C, and the NH_4OAc solution was changed every 8 h, then lyophilized to dryness and resuspended in 30 μL of 1 mM NH_4OAc and 30 μL of isopropanol. The adducted

RNA sample was analyzed by ESI^--MS to yield an experimental mass (3880.8) consistent with a monobisulfite adducted RNA strand (calcd = 3880.3).

Structural Analysis of the Monobisulfite Adduct to the Pseudouridine Nucleoside. All chemicals were obtained from commercial suppliers. The $NaHSO_3$ solution was freshly prepared before reaction as a 5 M stock solution (pH 5.0). The pseudouridine nucleoside (20 mM) was allowed to react with 3 M $NaHSO_3$ at pH 5.0 for 16 h at 50 °C to give a product yield of ∼90%. The reaction was analyzed using a Hypercarb HPLC column running a mobile phase combination of A = 20 mM NH_4OAc (pH 7) and B = MeOH with a flow rate = 1 mL/min while monitoring the elution profile at 220 and 260 nm. The method was held at 0% B for 5 min, after which B was changed to 95% B over 20 min via a linear gradient. The two product peaks were collected, dried, and submitted to mass spectrometric analysis in which they gave masses consistent with monoadducts of bisulfite to the nucleoside (calcd mass $[M-H]^-$ = 325.27 and expt mass for the first isomer = 325.00 and the second isomer = 325.07). The purified samples were analyzed by UV-vis in ddH_2O buffered with 20 mM NaP_i showing the first eluting peak named isomer 1 to have a λ_{max} = 265 nm and the second eluting peak named isomer 2 to have a λ_{max} = 266 nm. In a final experiment, the purified compounds were analyzed by ^{1}H-NMR: isomer 1 (500 MHz, D_2O) δ 7.60 (s, 1 H), 4.40 (s, 0 Hz, 1 H), 4.31 (dd, 8.32 Hz, 1 H), 3.67 (m, 2.45 Hz, 1 H), 3.61 (dd, 2.17 Hz, 1 H), 3.45 (dd, 7.33, 6.85, and 4.40 Hz, 1 H), and 3.22 (dd, 4.89 and 2.93 Hz, 1 H) and isomer 2 (500 MHz, D_2O) δ 7.61 (s, 1 H), 4.39 (d, 6.85 Hz, 1 H), 3.98 (t, 6.85 and 6.35 Hz, 1 H), 3.75 (t, 5.87 Hz, 1 H), 3.65 (dd, 3.91 and 2.93 Hz, 1 H), 3.58 (dd, 3.42 and 2.94 Hz, 1 H), and 3.44 (dd, 6.85 and 4.89 Hz, 1 H).

ACKNOWLEDGMENTS. We thank Brian Dalley in the Huntsman Cancer Institute high-throughput sequencing core for expert assistance, Danesh Moazed (Harvard Medical School) for support of V.K. during late stages, and Brenda Bass (University of Utah) for helpful comments on the manuscript. This work was supported by Grant R21 HG008058 from NIH/National Human Genome Research Institute, Howard Hughes Medical Institute (B.R.C.), and Grant P30 CA042014 from the National Cancer Institute for core facilities.

1. Machnicka MA, et al. (2013) MODOMICS: A database of RNA modification pathways–2013 update. *Nucleic Acids Res* 41:D262–D267.
2. Frye M, Jaffrey SR, Pan T, Rechavi G, Suzuki T (2016) RNA modifications: What have we learned and where are we headed? *Nat Rev Genet* 17:365–372.
3. Gilbert WV, Bell TA, Schaening C (2016) Messenger RNA modifications: Form, distribution, and function. *Science* 352:1408–1412.
4. Wang X, He C (2014) Dynamic RNA modifications in posttranscriptional regulation. *Mol Cell* 56:5–12.
5. Meyer KD, et al. (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* 149:1635–1646.
6. Dominissini D, et al. (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485:201–206.
7. Carlile TM, et al. (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* 515:143–146.
8. Dominissini D, et al. (2016) The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature* 530:441–446.
9. Edelheit S, Schwartz S, Mumbach MR, Wurtzel O, Sorek R (2013) Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genet* 9:e1003602.
10. Li X, et al. (2016) Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. *Nat Chem Biol* 12:311–316.
11. Li X, et al. (2015) Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat Chem Biol* 11:592–597.
12. Lovejoy AF, Riordan DP, Brown PO (2014) Transcriptome-wide mapping of pseudouridines: Pseudouridine synthases modify specific mRNAs in S. cerevisiae. *PLoS One* 9:e110799.
13. Schwartz S, et al. (2014) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* 159:148–162.
14. Squires JE, et al. (2012) Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res* 40:5023–5033.
15. Schaefer M (2015) RNA 5-methylcytosine analysis by bisulfite sequencing. *Methods Enzymol* 560:297–329.
16. Zheng G, et al. (2015) Efficient and quantitative high-throughput tRNA sequencing. *Nat Methods* 12:835–837.
17. Behm-Ansmant I, Helm M, Motorin Y (2011) Use of specific chemical reagents for detection of modified nucleotides in RNA. *J Nucleic Acids* 2011:408053.
18. Cozen AE, et al. (2015) ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nat Methods* 12:879–884.
19. Lei Z, Yi C (2017) A radiolabeling-free, qPCR-based method for locus-specific pseudouridine detection. *Angew Chem Int Ed Engl* 56:14878–14882.
20. Zaringhalam M, Papavasiliou FN (2016) Pseudouridylation meets next-generation sequencing. *Methods* 107:63–72.
21. Li Y, Tollefsbol TO (2011) DNA methylation detection: Bisulfite genomic sequencing analysis. *Methods Mol Biol* 791:11–21.
22. Khoddami V, et al. (2019) Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. Gene Expression Omnibus. Available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE90963. Deposited March 8, 2019.
23. Mosbruger TL (2019) RBSSeqTools. GitHub. Available at https://github.com/HuntsmanCancerInstitute/RBSSeqTools. Deposited September 29, 2016.
24. Legrand C, et al. (2017) Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. *Genome Res* 27:1589–1596.
25. Hauenschild R, et al. (2015) The reverse transcription signature of N-1-methyladenosine in RNA-Seq is sequence dependent. *Nucleic Acids Res* 43:9950–9964.
26. Macon JB, Wolfenden R (1968) 1-Methyladenosine. Dimroth rearrangement and reversible reduction. *Biochemistry* 7:3453–3458.
27. Li X, et al. (2017) Base-resolution mapping reveals distinct m(1)A methylome in nuclear- and mitochondrial-encoded transcripts. *Mol Cell* 68:993–1005 e9.
28. Safra M, et al. (2017) The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* 551:251–255.
29. Grozhik AV, Jaffrey SR (2017) Epitranscriptomics: Shrinking maps of RNA modifications. *Nature* 551:174–176.
30. Hoareau-Aveilla C, Henry Y, Leblanc T (2008) [Dyskeratosis congenita, a disease caused by defective telomere maintenance]. *Med Sci (Paris)* 24:390–398.
31. Townsley DM, Dumitriu B, Young NS (2014) Bone marrow failure and the telomeropathies. *Blood* 124:2775–2783.
32. Kiss T, Fayet-Lebaron E, Jády BE (2010) Box H/ACA small ribonucleoproteins. *Mol Cell* 37:597–606.
33. Autexier C, Lue NF (2006) The structure and function of telomerase reverse transcriptase. *Annu Rev Biochem* 75:493–517.
34. Mitchell JR, Wood E, Collins K (1999) A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* 402:551–555.
35. Kim NK, Theimer CA, Mitchell JR, Collins K, Feigon J (2010) Effect of pseudouridylation on the structure and activity of the catalytically essential P6.1 hairpin in human telomerase RNA. *Nucleic Acids Res* 38:6746–6756.
36. Singhal RP (1974) Chemical probe of structure and function of transfer ribonucleic acids. *Biochemistry* 13:2924–2932.
37. Shibutani S, Takeshita M, Grollman AP (1997) Translesional synthesis on DNA templates containing a single abasic site. A mechanistic study of the "A rule". *J Biol Chem* 272:13916–13922.