

http://pubs.acs.org/journal/acscii

Nanopore Dwell Time Analysis Permits Sequencing and **Conformational Assignment of Pseudouridine in SARS-CoV-2**

Aaron M. Fleming,* Nicole J. Mathewson, Shereen A. Howpay Manage, and Cynthia J. Burrows*



differences during sequencing in the base-calling, ionic current, and dwell time in the nanopore sensor; however, the signals were found to have a dependency on the context that would result in biases when sequencing unknown samples. A solution to the challenge was the identification that the passage of Ψ through the helicase brake produced a long-range dwell time impact with less context bias that was used for modification identification. The data analysis approach was employed to analyze publicly available direct RNA sequencing data for SARS-CoV-2 RNA taken from cell culture to



locate five conserved Ψ sites in the genome. Two sites were found to be substrates for pseudouridine synthase 1 and 7 in an in vitro assay, providing validation of the analysis. Utilization of the helicase as an additional sensor in direct RNA nanopore sequencing provides greater confidence in calling RNA modifications.

■ INTRODUCTION

The epitranscriptome represents the collection of chemical modifications on mRNA strands that are essential for their biological functions in cells. Complete nuclease digestion and mass spectrometric analysis of cellular RNAs from many organism types have identified >170 different modifications, a large subset of which have been found in humans.¹⁻⁴ One drawback to this analytical approach is the fact that sequence information regarding the sites modified, the specific RNA, and the extent of modification at a given site is lost. This information is essential to address how the epitranscriptome impacts RNA physiology. Methods developed for sequencing target modifications include high accuracy but low-throughput approaches such as SCARLET.⁵ At present, high-throughput methods relying on modification enrichment and nextgeneration sequencing (NGS), best illustrated by the many N^6 -methyladenosine (m⁶A) sequencing approaches reported, have enabled numerous discoveries of the m⁶A epitranscriptome.^{3,6} The development of techniques to sequence other RNA modifications with high accuracy and in a quantitative fashion is desperately needed to continue our exploration of the epitranscriptome.

Pseudouridine (Ψ), the most abundant global RNA modification, is an isomerization product of uridine (U) found at high relative levels (>1%) in eukaryotic tRNA and rRNA, as well as in viral RNA, and lower levels (<1%) in eukaryotic mRNA (Figure 1A).^{1,7,8} This modification is tied to

critical RNA functions in cells such as reinforcing RNA secondary structure and regulation of translation, and Ψ levels change in response to oxidative, micronutrient, or heat-shock stress.^{7,9–11} Initially, Ψ was located in RNA via digestion and chromatographic methods followed by mass spectrometric quantification approaches,^{12–14} and recently, high-throughput NGS using Ψ -specific chemistry has revealed sites in the mammalian transcriptome for Ψ . Pseudouridine is specifically alkylated by the carbodiimide CMC (N-cyclohexyl-N'-(2morpholinoethyl)carbodiimide metho-p-toluenesulfonate) to yield a stable and bulky adduct to stall reverse transcription, which is found by distinct sequencing read stops in comparison to a nonalkylated matched control.^{11,15–17} Alternatively, Ψ can be sequenced by the absence of reactivity with hydrazine, while the parent U readily reacts to yield strand breaks detected during NGS.¹⁸ As a result of analyzing sequencing stops, these approaches cannot achieve read-through in order to sequence multiple sites simultaneously in single strands; additionally, RNA structure-induced stops can yield false positives, and

Received: July 1, 2021 Published: September 15, 2021





© 2021 The Authors. Published by American Chemical Society



Figure 1. Direct RNA sequencing for Ψ by monitoring current vs time traces in a protein nanopore-helicase platform. (A) Isomerization of U yields Ψ . (B) Structural depiction of the CsgG-helicase nanopore setup used in the R9.1.4 MinION/Flongle flow cells manufactured by ONT. (C) Example ion current vs time trace. This figure was made using the PDB files 4UV3²³ and 2P6R²⁴ that were selected on the basis of the description of this system in the literature and a patent.^{25,26}

quantification of the modification is challenging to conduct accurately using these approaches.^{19,20} One chemical solution is to treat the suspect RNA with bisulfite to yield a stable sugar adduct on the Ψ nucleotide,^{21,22} which induces a deletion signature during cDNA synthesis, allowing more than one modification per strand to be sequenced. However, this approach does not yield quantitative deletions, and therefore, the extent of modification at suspected sites is challenging to obtain.

Third-generation sequencing (TGS) technologies utilizing the Oxford Nanopore Technologies' (ONT) or the Pacific BioSciences' waveguide platforms sequence single molecules of DNA or RNA that allow modifications to be directly observed.²⁷ With the direct sequencing of DNA or RNA, steps that introduce biases in the workflow can be omitted, such as the need for high-yielding and selective chemistry, reverse transcription, and/or PCR. The ONT or MinION device sequences RNA by ratcheting the strand 3' to 5' via a helicase motor protein through a protein nanopore sensor

under an electrophoretic force (Figure 1B). The current modulates as the strand passes through the nanopore sensor in a characteristic way to call the individual nucleotides using a trained neural network algorithm (Figure 1C). Because the signal from the sensor results from molecular interactions between the RNA and the protein nanopore, it is anticipated that epitranscriptomic modifications will interact differently to generate unique signatures for their identification. Direct nanopore sequencing applied to synthetic RNAs with 100% modification present (e.g., m⁶A, Ψ , N⁷-methylguanosine (m⁷G), 5-hydroxymethylcytidine, etc.) has demonstrated the feasibility of this approach.^{28,29} Sequencing of the 16s rRNA from *E. coli* with the MinION successfully called m⁷G and Ψ at known sites;³⁰ additionally, m⁶A and Ψ have been found by base-calling errors in tRNA at known sites, synthetic mRNA, and mRNA from cellular sources.^{29,31–34} While many Ψ sites were identified in the prior studies, our analysis revealed that many sites would be missed on samples with Ψ at unknown locations. We overcame this issue by adding dwell time analysis centered on the pausing of Ψ in the helicase brake such that, when combined with base-calling errors, Ψ could be reliably found in any sequence context.

We then used the helicase dwell differences in conjunction with the other data analysis approaches to inspect publicly available nanopore sequencing data for the SARS-CoV-2 RNA subgenomes for Ψ .³⁵ The data analysis converged on five conserved sites with high confidence at which Ψ resides in the viral RNA subgenomes near the 3' end. Lastly, small RNAs were synthesized with two of these suspected Ψ sites from SARS-CoV-2 centrally located; it was found that one is a substrate for pseudouridine synthase 1 and 7 (PUS1 and PUS7) and the other is a PUS1 substrate, providing biochemical validation of the sequencing data analysis. The results can aid future cell studies in the identification of in cellulo writers of Ψ in the SARS-CoV-2 RNA genome. These findings demonstrate that nanopore sequencing has a great potential to enable complete and quantitative sequencing of the epitranscriptome when dwell time analysis is included.

RESULTS

The CsgG protein nanopore in the R9.4.1 sequencing flow cells has a 5-nt sensing zone (i.e., k-mer) for RNA;³⁶ thus, the



Figure 2. Base-calling frequencies from direct RNA nanopore data when U or Ψ is present in biologically relevant sequence contexts. The bases are called for U or Ψ in (A) singly-, (B) doubly-, or (C) triply-modified contexts. The raw nanopore reads were base-called with Guppy v.4.0 followed by Minimap2 alignment, Samtools processing, and IGV visualization ($n \sim 24000$).³⁹⁻⁴¹

duplex DNA templates utilized for RNA synthesis by in vitro transcription (IVT) were designed to install UTP or Ψ TP centrally within a 5-nt window composed of known pseudouridinylation sites in human rRNA, tRNA, and mRNA that only had one U nucleotide in the window (Figure S1).^{37,38} To study two different Ψ contexts that include a U, shorter RNAs were synthesized by solid-phase synthesis and then sequenced. Unique to the present studies is that the Ψ sites were separated by >25-nts to allow their individual study as they moved through the sequencer top to bottom (helicase to nanopore); this would be the most common scenario in which Ψ would be sequenced from biological samples except where they are clustered as described below. This approach contrasts with those that synthesize model mRNAs with all U sites converted to Ψ .^{29,34} In the sequence contexts bearing doubly- or triply-modified sites, 2-nt of the biologically adjacent sequence were maintained on either side of the modified region so that the k-mer represented a native sequence (Figure S1). Different RNA strands were sequenced with some redundancy in the contexts to test the reproducibility of the sequencer from one strand to the next as well as at different positions relative to the ends of the RNA. After library preparation using the direct RNA sequence kit from ONT, the samples were sequenced on either a MinION or Flongle flow cell using CsgG nanopores (R.9.1.4) to achieve ~300 000 or ~30 000 reads, respectively, and the fast5 reads were base-called using the Guppy tool to yield fastq files for further analysis (Figure S2).

The sequencing reads were aligned to the reference using Minimap2 to identify the base-call identities for the 20 different sequence contexts studied. For the RNAs containing U, the called nucleotide was >90% U with the remainder called as C or A with sequence context dependency (Figure 2A). When Ψ was present, the nucleotide called was predominantly a mixture of C and U and a low amount of A or G, consistent with prior results.^{28–30,33,34} The new finding herein is that the distribution of C and U called at the 15 singly-modified Ψ sites ranged from 10% to 97% C with the remainder predominantly called as U, a false negative signal (Figure 2A). Inspection of the sequence-dependent base-calling results for Ψ did not lead to an obvious sequence context trend (Figure 2A). The demonstration of reproducibility in the base-calls was achieved by sequence redundancy in the strand design to identify similar base-calls for Ψ (Figure S3). The doubly- and triply-modified sites produced similar U vs C base-call signatures as the singlymodified sites, as well as impacted base-calling on the adjacent canonical nucleotides (Figure 2B,C). These data indicate that the Guppy algorithm, which was trained on canonical RNA nucleotides, when confronted with Ψ called the site as a C or U with dependency on the sequence context.

Base-calling errors provide a convenient approach to locate RNA modifications.^{29,34} We selected the Eligos2 tool to analyze the sequencing data by sample comparison of the Ψ strand with the U strand to inspect for differences in the error of specific bases (ESB) (Figures 3A and S4).²⁹ Radar plots illustrate that the ESB values increase at Ψ sites and are impacted by adjacent nucleotides (Figure 3A). The ESB values allow one to calculate an odds ratio (oddR) or *P*-value of statistical significance for the presence of modifications (Figures 3B and S4). The application of Eligos2 to locate Ψ when compared to a U-containing RNA of the same sequence identified observable oddR values at the modified sites (Figure 3B), consistent with the literature.⁴² The oddR values ranged



Figure 3. Eligos2 analysis to locate Ψ in direct RNA sequencing data shows sequence context dependency in the magnitude of the oddR values. (A) Two example radar plots of ESB values for singly-modified Ψ versus U sites in RNA. (B) A plot of oddR values for the Ψ -modified sites in the RNAs studied. More example ESB plots and the reproducibility plots for the oddR values found for Ψ are provided in Figures S4 and S5.

from 4 to 240 for the singly Ψ -modified sites with dependency on the sequence context (Figure 3B). A trend was observed in which those Ψ sites called to a greater extent as C gave the higher oddR values (Figures 2A and 3B). The doubly- and triply-modified sites were also detectable by the Eligos2 tool (Figure S4). A comparison of the redundant Ψ -modified contexts in the RNAs studied found Eligos2 can yield similar statistical values for some of the modified sites selected for study, and others were found to vary considerably between the test cases (Figure S5). Using base-calling errors provides a means to locate Ψ in a sequence, but the corresponding signals are sequence-context dependent that favorably bias detection of Ψ to sequence contexts that yield greater C base-calls and base-calling error.

The ONT single-molecule sequencing platform reads the nucleotide sequence by using an electrophoretic force and a helicase brake to slowly move the strand through a small aperture protein pore.²⁵ As the nucleotides pass through the constriction of the nanopore that is ~5-nt long (i.e., k-mer) for RNA,³⁶ the ionic current is deflected with dependency on the sequence identity inside the aperture. Modified nucleotides have different sizes, shapes, and/or hydrodynamic properties that permit changes in the current and/or dwell times to be detected compared to the canonical forms resulting in their identification. Thus, the raw nanopore data can be analyzed for the presence of RNA modifications.

Interrogation of the current intensity and dwell times requires one to align the base-called data to the events using either Tombo or Nanopolish to inspect the population of single-molecule reads at each point (Figures S6 and S7).^{42,43} The Nanocompore³¹ tool takes in the Nanopolish event alignments and provides the ability to compare the populations on modified RNAs against a matched population void in the



Figure 4. Current intensity and dwell time analysis of RNA with U or Ψ passing through the protein nanopore sensor. (A) A representation of the helicase-nanopore sequencing setup to illustrate where the data are analyzed. Example (B) current histograms and (C) dwell time histograms for U or Ψ in the CsgG portion of the nanopore sensor, in which the distributions were analyzed with Nanocompore to identify sites of statistical differences between the populations by pairwise analysis using the Kolmogorov–Smirnov test. The *P*-values from the statistical test were –log transformed to visualize the results of the test by increasing the signal at those most different at each site based on (D) current and (E) dwell time. The analysis was conducted across 10-nt in which the modification could span the S-nt window of the protein nanopore sensor region. The plots were constructed from >800 data points obtained from Nanopolish extraction of the currents and dwell times from the raw fast5 data files. More example histograms can be found in Figure S7.

modification to conduct a pairwise analysis using the Kolmogorov–Smirnov test to report *P*-values of statistical differences. The *P*-values are –log transformed to visualize sites with a significant difference by an increased value. Thirteen of the biologically relevant Ψ -containing sequence contexts were analyzed for differences in signal when the modification passed through the protein nanopore sensor (Figure 4A). The single-molecule events were compiled to make histograms of the current intensities or dwell times for the U or Ψ samples at each sequence read frame (Figure 4B,C; top or blue = U, bottom or red = Ψ). The inspection of the

current histograms for U vs Ψ as the suspected site moved through the 5-nt window of the sensor for each sequence context identified significant sites based on the Nanocompore analysis (Figures 4D and S7). The key observation regarding the current-level differences is that Ψ impacts the signal, but the position in the 5-nt window at which the impact is most significant varies with the sequence context.

Analysis of the CsgG dwell time differences between U vs Ψ in the contexts studied found a similar observation; Ψ can slightly impact the dwell time compared to U in the nanopore sensor, but the position in the known 5-nt window at which



Figure 5. Passage of Ψ through the helicase active site impacts the read dwell time compared to U that permits the detection of the epitranscriptomic modification at a distal site. (A) Schematic of the helicase-protein nanopore setup. Example (B) current and (C) dwell time histograms for a U or Ψ in the active site of the helicase. Stacked plots of $-\log(P$ -values) from the Nanocompore analysis for the complete passage of the suspect sites through the nanopore setup looking at statistical differences in the (D) current and (E) dwell times. (F) The median dwell times for the site most statistically significant based on Nanocompore analysis when it resides in the helicase. (G) Plot of the areas for the median dwell time distributions. The analyses were conducted on >1000 single-molecule measurements. Additional data are provided in Figure S9.

the difference is maximal is dependent on the context (Figure 4E). Because of this variability in the maximal difference in signal, the resolution to call the modified site in an unknown

sequence is 9-nt (i.e., k-mer = 5-nt window flanking both sides of a centrally located modification). The sequenced and analyzed redundant sequence contexts were compared to



Figure 6. Analysis of nanopore sequencing reads for the SARS-CoV-2 RNA subgenomes for Ψ . (A) Interrogation of the SARS-CoV-2 RNA extracted from cell culture with modifications against an IVT-generated genome without modifications to find statistically significant differences in current intensity and base-calling error for the TRS-S subgenome. (B) As an example, U28 759 in TRS-3a yields a base-calling error found by Eligos2 and a long-range dwell signature found by Nanocompore/Nanopolish analysis. (C) Plot illustrating the Ψ sites found in the analysis. (D) RNAfold prediction of the region flanking U28 927 and U29 418 to illustrate the local secondary structure. Data for the other subgenomes is provided in Figures S11–20, and the full RNAfold analysis is found in Figure S21.

evaluate the reproducibility of the $-\log$ transformed *P*-values; it was found that they were poorly reproducible (Figure S8). This is not surprising because of sampling errors leading to different levels of statistical significance.

During the inspection of the differences in currents and dwell times between the RNAs with U or Ψ , a long-range change to the dwell time population analysis was observed 10–11-nts 3' to the suspected sites. Sequencing RNA on the ONT platform occurs in the 3' to 5' direction (Figure 5A),²⁵ and therefore, we propose the long-range difference found only in the dwell time analysis is a result of the modification impacting the helicase activity (Figure 5B,C). The 10–11-nt registry

difference between the helicase active site and the nanopore sensor is supported by a similar observation previously reported between 10- and 12-nt.⁴⁴ More specifically, the current analysis at positions 10–11 3' to the modification site when the U/ Ψ site is in the helicase did not significantly change on the basis of Nanocompore analysis (Figure 5D); however, the dwell times produced a long-range signature for sequencing Ψ observed for each sequence context (Figure 5E black arrow).

The dwell time distributions for helicase stalling at the maximally different sites when plotted on a log axis were found to be Gaussian distributed for the U populations and bimodal

Gaussian distributed for the Ψ populations (Psi-1 and Psi-2; Figure 5B-E). The distributions were fit to Gaussian equations $(r^2 > 0.95)$, allowing one to determine the average event time for the populations (Figure 5B,C). First, the U sites were processed by the helicase with average event times in the range of 6-13 ms (Figure 5F blue bars). The event time subpopulation labeled Psi-1 gave an average time similar to the U population (Figure 5F black bars or Psi-1). In contrast, the second subpopulation for Ψ labeled Psi-2 produced a longer average time of 10-60 ms (Figure 5F red bars or Psi-2). The Psi-2 subpopulations had a >3-fold longer average time than the Psi-1 subpopulations for the same events. This observation found the helicase activity on Ψ as a substrate is split into two populations that result from the modified nucleotide interacting with the active site of the helicase in likely two different conformations.

The two Ψ helicase activity subpopulations were then integrated to determine whether the populations change as a function of sequence context (Figure 5G). The analysis found 5'-G Ψ sequence contexts gave >70% of the total area as the Psi-2 subpopulation had a larger average dwell time (Figure 5G). In contrast, the Psi-2 subpopulation was present with <30% of the events for the sequence context that had any other base 5' to Ψ besides G (Figure 5G). The Ψ impact on the dwell time while residing in the helicase active site was observed consistently in the redundant sequence contexts studied (Figure S10). The ability to use dwell time differences when Ψ passes through the helicase active site provides another means for modification sequencing that does not rely on the protein nanopore sensor.

The dwell time signature found for Ψ provides two advancements for sequencing modifications using the nanopore setup: (1) it minimizes the window to call a modification down to 2-nts, and (2) it provides a secondary approach to call modifications that is not reliant on error-prone signals from the protein nanopore sensor that were found to be sequencecontext dependent (Figures 2 and 3). The long-range dwell analysis was always present for Ψ , and when combined with base-calling error analysis, it results in greater confidence to call modifications and their locations in the nanopore sequencing data. Lastly, long-range interactions observed in the nanopore sensor when the modification passes through the motor protein have been noted but never applied to this extent.⁴⁴⁻⁴⁶

Nanopore sequencing of RNAs with U or Ψ within biologically relevant contexts found differences in the bases called (Figure 2), differences in base-calling errors (Figure 3), differences in the currents and dwell times when the sites resided in the nanopore sensor (Figure 4), and differences in the dwell times when the sites passed through the active site of the helicase (Figure 5). Only the long-range dwell time differences, those generated by the helicase, provide the ability to detect Ψ consistently in all sequence contexts, albeit with a weaker signal. We propose using a hybrid analytical approach of base-calling errors from nanopore sensor-derived signals using Eligos2 analysis coupled with dwell time signatures derived from helicase activity differences using Nanocompore/ Nanopolish analysis to permit high confidence calling of Ψ in RNAs.

This proposed approach was applied to publicly available nanopore sequencing data deposited for the SARS-CoV-2 RNA genome.^{35,47} In the deposited data, the modified RNA was obtained from SARS-CoV-2 infected cells, and the

nonmodified matched control was generated by IVT.35,47 The base-called fastq files were analyzed with Eligos2 to locate sites of base-calling errors between the samples, and the fast5 files were analyzed with Nanocompore/Nanopolish to inspect for electrical current and dwell time differences in the nanopore sensor and helicase proteins. A noteworthy point about coronaviruses is that, during replication of their genomes, large populations of subgenomic RNAs (sgRNAs) are generated, in which these shorter RNAs code for conserved structural and accessory proteins (spike protein [S], envelope protein [E], membrane protein [M], and nucleocapsid protein [N]) as well as for key accessory proteins (3a, 6, 7a, 7b, 8, and 10).³⁵ The analysis described in the text to locate Ψ inspected the sgRNAs or transcriptional regulatory sequences (TRSs) for the structural proteins S (TRS-S), M (TRS-M), and E (TRS-E) as well as the accessory protein 3a (TRS-3a), which are the longest of the population of sgRNAs. One more point is that homopolymer runs can yield signals that masquerade as modifications.^{29,48} The SARS-CoV-2 analysis described removed homopolymer runs >4-nts; these sites may be modified³⁵ but were removed because of the known issues with the sequencer that were verified in the RNAs studied (Figure S11).

Using TRS-S (length = 8407-nt) as an example, the basecalling error analysis to report oddR values identified 111 sites with a value ≥ 3 based on the data herein (Figures 3, 6A, and S12). The current analysis of TRS-S found 810 sites with a $-\log(P$ -value) threshold \geq 50 and 125 sites with a threshold set to ≥ 100 , a value selected to match the number of modified sites found with the base-calling error analysis (Figures 6A and S12). The nanopore data analysis approach described in the present work found five high confidence Ψ sites in TRS-S, five in TRS-3a, five in TRS-E, and five in TRS-M (Figures 6C red labels and S12-S15). In TRS-3a, -E, and -M, five of the identified peaks were at the same location (U27 164, U28 039, U28 759, U28 927, and U29 18); therefore, TRS-S was inspected at those positions to find weak base-calling error signals (oddR < 3) and weak long-range dwell times for Ψ that support modification at all of these sites except U28 039 with lower confidence (Figure 6C gray labels). The analysis found U nucleotides in the SARS-CoV-2 TRSs that are modified to Ψ with conservation through the subgenomic RNAs.

The knowledge of these five conserved sites of Ψ led us to inspect the other TRSs in SARS-CoV-2 to determine whether their occupancy spanned all subgenomes. The subgenomes progressively decrease in length, and therefore, if the sequence had the conserved Ψ site, it was found to be modified in every TRS (Figures S16-S20). The sequence contexts for the five identified sites were U27 164 = 5'-AGXGA, U28 039 = 5'-AGXAG, U28 759 = 5'-CCXCA, U28 927 = 5'-GCXCU, and U29 418 = 5'-CUXAC (Figure S21); in the first two sites of the list, they have the 5'-G Ψ context that likely gave strong dwell time signatures to permit their detection, and the other three had a 5'-pyrimidine that in general produced stronger base-calling errors (Figure 3). To reiterate, the analysis of the SARS-CoV-2 RNA sequencing data using nanopores identified five conserved pseudouridinylation sites on the 3' end of the genome fragments. The base-calling error and dwell time analysis hybrid approach enabled this discovery out of the noisy nanopore data. The hybrid analysis approach did find other nucleotides that could be modified, but they were not explored further because standards for these were not studied herein (Figures S11-S20); although there are examples of possible m⁶A modification sites found that were reported by commonly applied sequencing methods for this modification (A27 525, A29 428, and A29 658,⁴⁹ and a few that were not previously reported at A24 420 in TRS-S and A27 334 in TRS-3a), it is not known if these are bonified modification sites.

Sequences of 50 nucleotides flanking each of the five conserved sites were submitted for RNAfold analysis, and it was found that three had the suspect site in hairpins at the base of the stem, near a bulge, or near a loop (Figures 6 and S21). These structures could be possible PUS1 or PUS7 substrates;³⁸ accordingly, we made recombinant PUS1 and PUS7 and allowed them to react on small synthetic RNAs with U28 927 or U29 418 centrally located. The determination of the presence of Ψ was achieved by a literature gel-based protocol.⁵⁰ The gel analysis found U28 927 is both a PUS1 and a PUS7 substrate, while U29 418 is a PUS1 substrate in vitro (Figure S22). These observations provide some biochemical validation for the data analysis for Ψ in the nanopore sequencing reads from SARS-CoV-2.

DISCUSSION

Direct RNA sequencing with nanopores has the potential to locate epitranscriptomic modifications via current levels, dwell times, and the associated base-calling errors. At present, the signatures for all modifications are not known and studies are needed to address what the signals are as well as what the biases and limitations are to the data analysis. In the present work, Ψ was synthetically incorporated by IVT in RNA at known locations in 18 different human-relevant sequence contexts found in rRNA, mRNA, and tRNA (Figure S1). The Ψ sites were spaced >25-nt apart to study them one at a time as they pass from the helicase to the nanopore sensor, an overall distance that spans \sim 17-nt from the entry of the helicase to the exit of the k-mer sensing zone in the protein nanopore. Pseudouridine is base-called by Guppy predom-^{8;34} and the inantly as U or C, consistent with other studies,² present work found the ratio is dependent on the local sequence context (Figure 2). The base-calling error for Ψ was greater than U permitting detection of the modification via Eligos2 (Figure 3); however, the base-calling errors were sequence context dependent, similar to the base-calling differences. Two extreme examples found in the data illustrate the challenges in using base-calling data alone for the RNA modification sequence; in the 5'-ACXCA (X = U or Ψ) context, the U/C ratio is 7:88 with a base-calling error analysis giving an oddR value of 233, while the similar 5'-CAXCG context had a U/C ratio of 90:10 and an oddR value of 4, both when 100% Ψ is present at position X (Figures 2 and 3). In real samples, this approach will systematically favor observation of high error and high C calling sites over those that fit a profile similar to the reactant U, resulting in huge biases to the data, especially at sites that are not quantitatively modified. This is a challenge because Ψ can reside in all possible sequence contexts.¹

A similar sequence dependency was observed for detecting Ψ when inspecting the raw currents and dwell times as the modification passed through the protein nanopore sensor (Figure 4); moreover, the position within the k-mer window for which Ψ impacted the current and/or dwell time to the greatest extent was dependent on the sequence context, resulting in a 9-nt ambiguity of the location of a modification in a real sample. These analytical approaches work for modifications like N^6 -methyladenosine that are favorably

deposited in reproducible sequence contexts^{3,6} but fail for Ψ that can exist in all possible sequence contexts.³⁸

The observation that Ψ impacts the dwell time as it passes through the helicase active site compared to U alleviates some of the challenges for detection, especially when this analytical approach is used in tandem with other detection strategies, as we propose in the present work. In all sequence contexts, Ψ produced an observable signal not seen for U in the helicase, which slows the helicase processing activity by \geq 3-fold for a subpopulation of the events (Figure 5). The longer dwell time subpopulation distribution was greatest for 5'-G Ψ sequence contexts, leading these contexts to be the easiest in which to detect Ψ . Nonetheless, in all sequence contexts studied (Figure 5F), the signal was present, yielding a positive signal to locate Ψ in the strand. Unlike the other methods that did not report consistent values on replicate studies (Figures S3 and S5) and in the case of the nanopore sensor that did not yield signals with single nucleotide resolution (Figure 4D), the helicase stalling leading to a dwell time signature detects Ψ within a 2nt window created by the helicase active site (Figure 5E). With appropriate synthetic control RNA strands for sequence contexts to calibrate against, these data could provide quantitative information on the extent of epitranscriptomic modification at a suspected Ψ site. Recent reports using direct RNA nanopore sequencing to locate Ψ in a cellular transcriptome analyzed base-calling error; as described above, this approach is sequence-context biased. However, the inspection for base-calling errors computationally is straightforward and can be implemented on low sequence coverage data sets. The method outlined here is computationally demanding and requires higher sequence coverage data sets for analysis; thus, our approach is likely best applied for targeted RNA sequencing for Ψ in viral RNA genomes, rRNA, or tRNA.

Pseudouridine creates four differences in the RNA strand compared to U that likely led to the sequencing signatures observed: (1) Uridine has a hydrogen bond donor site at N3, while Ψ can hydrogen bond at N1 and N3, and both have the hydrogen bond acceptor sites at O^2 and O^4 (Figure 1A). The Ψ N1 hydrogen shows long-lived bonding with the phosphodiester backbone to introduce rigidity in duplex RNA, and the bond likely exists in single-stranded RNA albeit with a shorter lifetime.⁵¹ (2) The glycosidic bond angle for Ψ shows a slight syn preference while U adopts the anti conformation almost exclusively.⁵² (3) Pseudouridine is more hydrophilic than U, and (4) Ψ stacks with adjacent bases better than U.53 These physical differences likely result in the ability to differentiate Ψ from U in the nanopore sequencer. In the CsgG protein nanopore, calls of Ψ as U or C with sequence context dependency may result from the hydrogen bond and syn/anti conformational differences that impact the interactions with the nanopore. We propose the U vs C base-calling ratio is a result of the syn vs anti conformation of the Ψ heterocycle. The syn vs anti conformation distribution shows strong sequence context dependency as observed in the wide range of U/C base-calling ratios (Figure 2), and this provides a molecular understanding to the incoherent signal in the nanopore from this isomer of the U nucleotide.

As for the helicase, a patent suggests that a mutant form of a Hel308 helicase is used for RNA sequencing in the ONT platform.²⁶ Helicases such as Hel308 predominantly interact with nucleic acids via the backbone, although an amino acid wedge interacts with the base pair to be broken (-1 position) and the pair just broken (+1 position) during the unwinding

process,²⁴ assuming this amino acid was not mutated in the helicase used. A crystal structure for an archaeal Hel308 helicase identified that Phe and Arg residues π stack with the bases at the -1 and +1 positions, respectively.²⁴ The interesting observation is that this provides a two-nucleotide window in which the Ψ base could impact the helicase activity, consistent with the present findings for detection of the isomer via helicase stalling during sequencing. All four Ψ – U differences discussed may contribute to the helicase differentiation of Ψ , while the local rigidity imposed by N1–H of Ψ and its better π stacking are likely the dominant forces leading to the slower helicase processing kinetics. Further, the 5' G effect likely occurs from more stabilized π stacking that can compete with Phe and/or Arg to slow the helicase translocation along the RNA strand. Lastly, the reason why Ψ within the helicase active site results in two different average time populations is again a consequence of the syn/anti conformation distribution found for this heterocycle. This observation suggests one face of $\Psi \pi$ stacks better with Phe than the other face. Details of the helicase mutations are needed to better address the Ψ vs U differences that enabled differentiation of U and Ψ .

The helicase stalling at Ψ permitted analysis for this modification with greater confidence in the noisy sequencing reads for the SARS-CoV-2 RNA genome. In the subgenomic TRSs, five conserved Ψ sites on the 3' end of the RNA subgenomes exist (Figures 6C and S21). The structure of RNA guides where Ψ is installed by pseudouridine synthases (PUSs) that are stand-alone enzymes such as PUS1, PUS7, PUS7L, and TRUB1.¹⁵⁻¹⁷ Cells infected with RNA viruses, such as SARS-CoV-2, were found to have slight upregulation of many PUS enzymes that include PUS1 and PUS7.54 Each of the five sites with 50-nts of flanking native sequence were selected and submitted for RNAfold analysis to identify their predicted folding (Figures 6D and S21).55 The predicted folds for regions around U28 759, U28 927, and U29 418 place the U at the base of a hairpin or in a bulge, which are structures previously found to be PUS1 substrates (Figures 6D and S21).³⁸ As for the other two sites, U27 164 is in a large singlestranded loop and U28 039 is near the middle of a long duplex RNA adjacent to a G/U wobble base pair (Figure S21). In vitro studies with recombinant PUS1 and PUS7 showed that two SARS-CoV-2 sites were possible U substrates for Ψ installation (U28 927 and U29 418). A challenge with this in vitro analysis is that the enzymes were only given one substrate to be reacted upon that could yield site selection not found in cells, in which sites of modification are determined by many other factors not present in a test tube. Nonetheless, the biochemical studies provide some validation for the sequencing studies that found Ψ in the viral RNA. Studies on cells infected with SARS-CoV-2 while knocking down or out the PUSs are needed to truly define the Ψ writer(s) in the viral RNA.

Inspections of other RNA viruses (e.g., the flaviviruses Zika and HCV) by LC-MS have found Ψ exists at ~1–2% of the U nucleotides;⁸ the occupancy of Ψ in SARS-CoV-2 is not known at present but may exist at a similar level as in the flaviviruses, considering their similar replication cycle and RNA-based genomes. Selecting TRS-E as an example, the hybrid approach used to locate possible Ψ nucleotides found five that represent 0.5% of the U nucleotides in this sequence; in contrast, Eligos2 calls 3.7% of the U nucleotides as modified (oddR \geq 3), and Nanocompore found 6.1% of the modified kmers had a U nucleotide defined as $-\log(P$ -value) \geq 100 (Figure S13). The Nanocompore results report on 5-nt k-mers, so the high value likely exists because of other chemical modifications in the genome. The oddR approach of modification calling by Eligos2 implies there will be false positives in the data set, and therefore, the number of modified nucleotides will be inflated. The approach herein is most likely an underestimate but reports on those sites that give complementary positive signals, and therefore, these are sites most anticipated to be modified at high levels and have the greatest likelihood of biological significance.

High occupancy Ψ sites in the SARS-CoV-2 RNA genome may have a biological function. The pseudouridinylation of viral RNA is hypothesized to be a mechanism by which the virus hijacks host enzymes to avoid the immune response. Support for this is the use of Ψ or its N1-methyl derivative for this purpose in mRNA vaccines⁵⁶ as well as the possibility that Ψ serves the same function in HIV and flaviviruses.^{57,58} If Ψ is used by SARS-CoV-2 and possibly other coronaviruses during the infection cycle to minimize immune stimulation, this points to new schemes for intervention;⁵⁹ additionally, if Ψ is found to be essential for infectivity of the virus, an inspection of host genetic variants for pseudouridine synthase enzymes that impact activity may reveal more clues as to why host outcomes from SARS-CoV-2 infection are variable beyond other comorbidity factors.

CONCLUSIONS

The epitranscriptome is currently at center stage for the discovery of critical details regarding RNA regulation in biology and is being fueled by advancements in sequencing technology. A nanopore sequencing platform composed of two proteins, a nanopore sensor and a helicase brake, enabled us to directly sequence RNA for Ψ , which is the most common RNA modification. Two key sequencing features were identified: (1) The protein nanopore sensor produced a wide range of signals for Ψ with dependency on the sequence context and position in the 5-nt sensing zone of the CsgG protein (Figures 2, 3, and 4). Some contexts gave robust signals and others were nearly in the noise that would yield false negatives when inspecting unknown samples. (2) The helicase employed to regulate the speed of translocation was found to have sensing capabilities for Ψ by stalling on the modification but not the parent U. The stalling results in a long-range dwell signal 10-11-nt 3' to the protein nanopore as the modification passes through the helicase active site (Figure 5). This signal was always present but most pronounced in 5'-G Ψ sequence contexts. Knowledge of base-calling errors and helicase dwell time signatures permitted the analysis of the SARS-CoV-2 RNA subgenomes for Ψ (Figure 6). Analysis of the viral RNA identified five conserved Ψ s on the 3' end of the fragments. The local structures for three of the modified sites are similar to those previously identified as PUS1 sites (Figures 6D and S21).³⁸ Using recombinant PUS1 and PUS7 as catalysts for Ψ introduction in synthetic RNAs, we found U28 927 is a substrate for both enzymes and U29 418 is a substrate for PUS1, providing biochemical validation of the data analysis; however, the actual writer enzyme(s) in host-infected cells is/ are not known, and future cell-based studies are needed to identify the synthases. Using the literature as a guide, we propose these Ψ s are beneficial to the virus by aiding in avoidance of the immune response to favor replication.⁵ The findings herein expand our knowledge of the viral epitranscriptome regarding Ψ , which can be exploited for

future interventions and understanding of individual host responses to SARS-CoV-2 viral infection.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscentsci.1c00788.

Experimental section, RNA sequences studied, read statistics, sequencing reproducibility data, Eligos2 results, Tombo results, Nanocompore results, dwell time histograms, homopolymer sequencing data, SARS-CoV-2 data analysis, predicted RNA folds, and PUS assays (PDF)

AUTHOR INFORMATION

Corresponding Authors

Cynthia J. Burrows – Department of Chemistry, University of Utah, Salt Lake City, Utah 84112-0850, United States; orcid.org/0000-0001-7253-8529; Email: burrows@ chem.utah.edu

Aaron M. Fleming – Department of Chemistry, University of Utah, Salt Lake City, Utah 84112-0850, United States;
orcid.org/0000-0002-2000-0310; Email: afleming@ chem.utah.edu

Authors

Nicole J. Mathewson – Department of Chemistry, University of Utah, Salt Lake City, Utah 84112-0850, United States

Shereen A. Howpay Manage – Department of Chemistry, University of Utah, Salt Lake City, Utah 84112-0850, United States

Complete contact information is available at: https://pubs.acs.org/10.1021/acscentsci.1c00788

Notes

The authors declare the following competing financial interest(s): A.M.F. is an occasional consultant for Electronic BioSciences, although EBS technology is not being used here. The data are available upon request.

ACKNOWLEDGMENTS

The authors are thankful to Dr. Yun Ding and Ms. Andrea Taillacq for their assistance with the computational tools. The National Institute of Health provided financial support for this project (R01 GM093099).

REFERENCES

(1) Jones, J. D.; Monroe, J.; Koutmou, K. S. A molecular-level perspective on the frequency, distribution, and consequences of messenger RNA modifications. *Wiley Interdiscip. Rev.: RNA* **2020**, *11*, e1586.

(2) Roundtree, I. A.; Evans, M. E.; Pan, T.; He, C. Dynamic RNA modifications in gene expression regulation. *Cell* **2017**, *169*, 1187–1200.

(3) Linder, B.; Jaffrey, S. R. Discovering and mapping the modified nucleotides that comprise the epitranscriptome of mRNA. *Cold Spring Harbor Perspect. Biol.* **2019**, *11*, a032201.

(4) Song, J.; Yi, C. Chemical modifications to RNA: A new layer of gene expression regulation. *ACS Chem. Biol.* **201**7, *12*, 316–325.

(5) Liu, N.; Parisien, M.; Dai, Q.; Zheng, G.; He, C.; Pan, T. Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* **2013**, *19*, 1848–1856.

(6) Hsu, P. J.; He, C. Identifying the m(6)A methylome by affinity purification and sequencing. *Methods Mol. Biol.* **2018**, *1649*, 49–57. (7) Borchardt, E. K.; Martinez, N. M.; Gilbert, W. V. Regulation and function of RNA pseudouridylation in human cells. *Annu. Rev. Genet.* **2020**, *54*, 309–336.

(8) McIntyre, W.; Netzband, R.; Bonenfant, G.; Biegel, J. M.; Miller, C.; Fuchs, G.; Henderson, E.; Arra, M.; Canki, M.; Fabris, D.; Pager, C. T. Positive-sense RNA viruses reveal the complexity and dynamics of the cellular and viral epitranscriptomes during infection. *Nucleic Acids Res.* **2018**, *46*, 5776–5791.

(9) Sakakibara, Y.; Chow, C. S. Role of pseudouridine in structural rearrangements of helix 69 during bacterial ribosome assembly. *ACS Chem. Biol.* **2012**, *7*, 871–878.

(10) Eyler, D. E.; Franco, M. K.; Batool, Z.; Wu, M. Z.; Dubuke, M. L.; Dobosz-Bartoszek, M.; Jones, J. D.; Polikanov, Y. S.; Roy, B.; Koutmou, K. S. Pseudouridinylation of mRNA coding sequences alters translation. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 23068–23074.

(11) Li, X.; Zhu, P.; Ma, S.; Song, J.; Bai, J.; Sun, F.; Yi, C. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* **2015**, *11*, 592–597.

(12) Cohn, W. E.; Volkin, E. Nucleoside-5'-phosphates from ribonucleic acid. *Nature* 1951, 167, 483-484.

(13) Davis, F. F.; Allen, F. W. Ribonucleic acids from yeast wich contain a fifth nucleotide. J. Biol. Chem. 1957, 227, 907–915.

(14) Crain, P. F.; McCloskey, J. A. The RNA modification database. *Nucleic Acids Res.* **1996**, *24*, 98–99.

(15) Carlile, T. M.; Rojas-Duran, M. F.; Zinshteyn, B.; Shin, H.; Bartoli, K. M.; Gilbert, W. V. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **2014**, *515*, 143–146.

(16) Lovejoy, A. F.; Riordan, D. P.; Brown, P. O. Transcriptomewide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in *S. cerevisiae*. *PLoS One* **2014**, *9*, e110799.

(17) Schwartz, S.; Bernstein, D. A.; Mumbach, M. R.; Jovanovic, M.; Herbst, R. H.; Leon-Ricardo, B. X.; Engreitz, J. M.; Guttman, M.; Satija, R.; Lander, E. S.; Fink, G.; Regev, A. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **2014**, *159*, 148–162.

(18) Marchand, V.; Pichot, F.; Neybecker, P.; Ayadi, L.; Bourguignon-Igel, V.; Wacheul, L.; Lafontaine, D. L. J.; Pinzano, A.; Helm, M.; Motorin, Y. HydraPsiSeq: a method for systematic and quantitative mapping of pseudouridines in RNA. *Nucleic Acids Res.* **2020**, *48*, e110.

(19) Zaringhalam, M.; Papavasiliou, F. N. Pseudouridylation meets next-generation sequencing. *Methods* **2016**, *107*, 63–72.

(20) Xue, C.; Zhao, Y.; Li, L. Advances in RNA cytosine-5 methylation: detection, regulatory mechanisms, biological functions and links to cancer. *Biomark. Res.* **2020**, *8*, 43.

(21) Fleming, A. M.; Alenko, A.; Kitt, J. P.; Orendt, A. M.; Flynn, P. F.; Harris, J. M.; Burrows, C. J. Structural elucidation of bisulfite adducts to pseudouridine that result in deletion signatures during reverse transcription of RNA. *J. Am. Chem. Soc.* **2019**, *141*, 16450–16460.

(22) Khoddami, V.; Yerra, A.; Mosbruger, T. L.; Fleming, A. M.; Burrows, C. J.; Cairns, B. R. Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 6784–6789.

(23) Goyal, P.; Krasteva, P. V.; Van Gerven, N.; Gubellini, F.; Van den Broeck, I.; Troupiotis-Tsaïlaki, A.; Jonckheere, W.; Péhau-Arnaudet, G.; Pinkner, J. S.; Chapman, M. R.; Hultgren, S. J.; Howorka, S.; Fronzes, R.; Remaut, H. Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. *Nature* **2014**, *516*, 250–253.

(24) Büttner, K.; Nehring, S.; Hopfner, K. P. Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat. Struct. Mol. Biol.* **2007**, *14*, 647–652.

(25) Branton, D.; Deamer, D. Nanopore Sequencing: An Introduction; World Scientific Publishing Co. Pte. Ltd., 2019. (26) Heron, A.; Clarke, J.; Moysey, R.; Wallace, J.; Bruce, M.; Jayasinghe, L.; Caprotti, D.; Soeroes, S.; McNeill, L.; Alves, D.; Bowen, R.; Milton, J. Modified Helicases. US 2015/0191709 A1, July 9, 2015.

(27) Motorin, Y.; Marchand, V. Analysis of RNA modifications by second- and third-generation deep sequencing: 2020 update. *Genes* **2021**, *12*, 278.

(28) Ramasamy, S.; Sahayasheela, V. J.; Yu, Z.; Hidaka, T.; Cai, L.; Sugiyama, H.; Pandian, G. N. Chemical probe-based nanopore sequencing to selectively assess the RNA modifications. *bioRxiv* 2021; DOI: 10.1101/2020.05.19.105338.

(29) Jenjaroenpun, P.; Wongsurawat, T.; Wadley, T. D.; Wassenaar; Trudy, M.; Liu, J.; Dai, Q.; Wanchai, V.; Akel, N. S.; Jamshidi-Parsian, A.; Franco, A. T.; Boysen, G.; Jennings, M. L.; Ussery, D. W.; He, C.; Nookaew, I. Decoding the epitranscriptional landscape from native RNA sequences. *Nucleic Acids Res.* **2021**, *49*, e7.

(30) Smith, A. M.; Jain, M.; Mulroney, L.; Garalde, D. R.; Akeson, M. Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. *PLoS One* **2019**, *14*, e0216709.

(31) Leger, A.; Amaral, P. P.; Pandolfini, L.; Capitanchik, C.; Capraro, F.; Barbieri, I.; Migliori, V.; Luscombe, N. M.; Enright, A. J.; Tzelepis, K.; Ule, J.; Fitzgerald, T.; Birney, E.; Leonardi, T.; Kouzarides, T. RNA modifications detection by comparative nanopore direct RNA sequencing. *bioRxiv* **2019**; DOI: 10.1101/843136. (32) Liu, H.; Begik, O.; Lucas, M. C.; Ramirez, J. M.; Mason, C. E.;

Wiener, D.; Schwartz, S.; Mattick, J. S.; Smith, M. A.; Novoa, E. M. Accurate detection of m6A RNA modifications in native RNA sequences. *Nat. Commun.* **2019**, *10*, 4079.

(33) Thomas, N.; Poodari, V.; Jain, M.; Olsen, H.; Akeson, M.; Abu-Shumays, R. Direct nanopore sequencing of individual full length tRNA strands. *bioRxiv* **2021**; DOI: 10.1101/2021.1104.1126.441285.

(34) Begik, O.; Lucas, M. C.; Pryszcz, L. P.; Ramirez, J. M.; Medina, R.; Milenkovic, I.; Cruciani, S.; Liu, H.; Vieira, H. G. S.; Sas-Chen, A.; Mattick, J. S.; Schwartz, S.; Novoa, E. M. Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nat. Biotechnol.* **2021**, DOI: 10.1038/s41587-021-00915-6.

(35) Kim, D.; Lee, J. Y.; Yang, J. S.; Kim, J. W.; Kim, V. N.; Chang, H. The architecture of SARS-CoV-2 transcriptome. *Cell* **2020**, *181*, 914–921.

(36) Rang, F. J.; Kloosterman, W. P.; de Ridder, J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.* **2018**, *19*, 90.

(37) Taoka, M.; Nobe, Y.; Yamaki, Y.; Sato, K.; Ishikawa, H.; Izumikawa, K.; Yamauchi, Y.; Hirota, K.; Nakayama, H.; Takahashi, N.; Isobe, T. Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res.* **2018**, *46*, 9289–9298. (38) Carlile, T. M.; Martinez, N. M.; Schaening, C.; Su, A.; Bell, T.

A.; Zinshteyn, B.; Gilbert, W. V. mRNA structure determines modification by pseudouridine synthase 1. *Nat. Chem. Biol.* 2019, 15, 966–974.

(39) Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100.

(40) Danecek, P.; Bonfield, J. K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M. O.; Whitwham, A.; Keane, T.; McCarthy, S. A.; Davies, R. M.; Li, H. Twelve years of SAMtools and BCFtools. *GigaScience* **2021**, *10*, giab008.

(41) Robinson, J. T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E. S.; Getz, G.; Mesirov, J. P. Integrative genomics viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26.

(42) Stoiber, M.; Quick, J.; Egan, R.; Eun Lee, J.; Celniker, S.; Neely, R. K.; Loman, N.; Pennacchio, L. A.; Brown, J. De novo identification of DNA modifications enabled by genome-guided nanopore signal processing. *bioRxiv* **2017**; DOI: 10.1101/094672.

(43) Loman, N. J.; Quick, J.; Simpson, J. T. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods* **2015**, *12*, 733–735.

(44) Stephenson, W.; Razaghi, R.; Busan, S.; Weeks, K. M.; Timp, W.; Smibert, P. Direct detection of RNA modifications and structure using single molecule nanopore sequencing. *bioRxiv* 2020, DOI: 10.1101/2020.05.31.126763.

(45) Manrao, E. A.; Derrington, I. M.; Laszlo, A. H.; Langford, K. W.; Hopper, M. K.; Gillgren, N.; Pavlenok, M.; Niederweis, M.; Gundlach, J. H. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* **2012**, *30*, 349–353.

(46) Zhao, X.; Liu, Y.; Chen, X.; Mi, Z.; Li, W.; Wang, P.; Shan, X.; Lu, X. Detection and characterization of single cisplatin adducts on DNA by nanopore sequencing. *ACS Omega* **2021**, *6*, 17027–17034.

(47) Miladi, M.; Fuchs, J.; Maier, W.; Weigang, S.; Pedrosa, N. D. I.; Weiss, L.; Lother, A.; Nekrutenko, A.; Ruzsics, Z.; Panning, M.; Kochs, G.; Gilsbach, R.; Grüning, B. The landscape of SARS-CoV-2 RNA modifications. *bioRxiv* **2020**; DOI: 10.1101/ 2020.1107.1118.204362.

(48) Jain, M.; Koren, S.; Miga, K. H.; Quick, J.; Rand, A. C.; Sasani, T. A.; Tyson, J. R.; Beggs, A. D.; Dilthey, A. T.; Fiddes, I. T.; Malla, S.; Marriott, H.; Nieto, T.; O'Grady, J.; Olsen, H. E.; Pedersen, B. S.; Rhie, A.; Richardson, H.; Quinlan, A. R.; Snutch, T. P.; Tee, L.; Paten, B.; Phillippy, A. M.; Simpson, J. T.; Loman, N. J.; Loose, M. Nanopore sequencing and assembly of a human genome with ultralong reads. *Nat. Biotechnol.* **2018**, *36*, 338–345.

(49) Liu, J. e.; Xu, Y.-P.; Li, K.; Ye, Q.; Zhou, H.-Y.; Sun, H.; Li, X.; Yu, L.; Deng, Y.-Q.; Li, R.-T.; Cheng, M.-L.; He, B.; Zhou, J.; Li, X.-F.; Wu, A.; Yi, C.; Qin, C.-F. The m6A methylome of SARS-CoV-2 in host cells. *Cell Res.* **2021**, *31*, 404–414.

(50) Carlile, T. M.; Rojas-Duran, M. F.; Gilbert, W. V. Transcriptome-wide identification of pseudouridine modifications using pseudo-seq. *Curr. Protoc. Mol. Biol.* **2015**, *112*, 4.25.21.

(51) Kierzek, E.; Malgowska, M.; Lisowiec, J.; Turner, D. H.; Gdaniec, Z.; Kierzek, R. The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res.* **2014**, *42*, 3492–3501.

(52) Neumann, J. M.; Tran-Dinh, S.; Bernassau, J. M.; Gueron, M. Comparative conformations of uridine and pseudouridine and their derivatives. *Eur. J. Biochem.* **1980**, *108*, 457–463.

(53) Davis, D. R. Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res.* **1995**, 23, 5020–5026.

(54) Blanco-Melo, D.; Nilsson-Payant, B. E.; Liu, W.-C.; Uhl, S.; Hoagland, D.; Møller, R.; Jordan, T. X.; Oishi, K.; Panis, M.; Sachs, D.; Wang, T. T.; Schwartz, R. E.; Lim, J. K.; Albrecht, R. A.; tenOever, B. R. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* **2020**, *181*, 1036–1045.

(55) Gruber, A. R.; Lorenz, R.; Bernhart, S. H.; Neuböck, R.; Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Res.* **2008**, *36*, W70–74.

(56) Sahin, U.; Karikó, K.; Türeci, Ö. mRNA-based therapeutics — developing a new class of drugs. *Nat. Rev. Drug Discovery* **2014**, *13*, 759–780.

(57) Marceau, C. D.; Puschnik, A. S.; Majzoub, K.; Ooi, Y. S.; Brewer, S. M.; Fuchs, G.; Swaminathan, K.; Mata, M. A.; Elias, J. E.; Sarnow, P.; Carette, J. E. Genetic dissection of Flaviviridae host factors through genome-scale CRISPR screens. *Nature* **2016**, *535*, 159–163.

(58) Martinez Campos, C.; Tsai, K.; Courtney, D. G; Bogerd, H. P; Holley, C. L.; Cullen, B. R. Mapping of pseudouridine residues on cellular and viral transcripts using a novel antibody-based technique. *RNA* **2021**, rna.078940.121.

(59) Cross, R. Epitranscriptomics: The new RNA code and the race to drug it. *Chem. Eng. News* **2019**, *97*, 34.