Nucleic Acids Research, 2026, **54**, gkag349
https://doi.org/10.1093/nar/gkag349
Computational Biology

Systematic benchmarking of *dorado* basecalling models for RNA modification detection with highly multiplexed nanopore sequencing

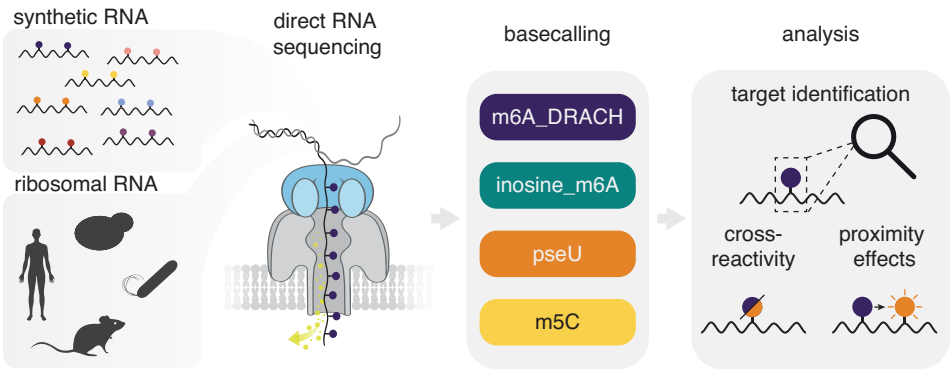
Gregor Diensthuber[1,2], Ivan Milenkovic[1], Laia Llovera[1], Ana Milovanovic[1], Francesco Pelizzari[1], Eva Maria Novoa[1,2,3,*]

[1]Center for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr Aiguader 88, Barcelona 08003, Spain
[2]Universitat Pompeu Fabra (UPF), Barcelona 08003, Spain
[3]ICREA, Passeig Lluís Companys 23, Barcelona 08010, Spain
*To whom correspondence should be addressed. Email: eva.novoa@crg.eu

Abstract

Nanopore direct RNA sequencing holds promise for advancing our understanding of the epitranscriptome. Recently, Oxford Nanopore Technologies released basecalling models capable of detecting *N*6-methyladenosine (m^6A), inosine (I), pseudouridine (Ψ), and 5-methylcytosine (m^5C). However, their performance and cross-reactivity with other modifications remain largely unexplored. Here, we systematically benchmark four available modification-aware basecalling models by evaluating their per-read and per-site predictions across synthetic molecules and biological samples from diverse species. Models performed well on highly modified, balanced synthetic constructs (AUC = 0.93–0.97, PR-AUC = 0.84–0.91), but their performance dropped sharply on unbalanced datasets that reflect modification abundances in biological samples (PR-AUC: 0.04–0.09). Analysis of *in vivo* rRNA samples confirmed this limitation, with false-discovery rate ranging from 50% to 100%, even after filtering with modification-free controls. We identify two major sources of false positives: cross-reactivities with other modifications and current alterations at sites neighbouring a modified residue. Finally, we demonstrate that basecalling error- and current-based methods can accurately detect modifications, offering effective alternatives for modifications lacking dedicated models. Our results highlight the utility and limitations of modification-aware basecalling models for RNA modification detection, and underscore the importance of including control samples to mitigate false-positive predictions.

Graphical abstract



Introduction

RNA modifications, collectively referred to as 'the epitranscriptome', are chemical moieties present on RNA molecules, fine-tuning their structure, function, and molecular fate [1, 2]. Recent advances in native RNA sequencing have enabled the direct detection of RNA modifications, bypassing the need for selective antibodies or chemical probing reagents [3–5]. Notably, nanopore direct RNA sequencing (DRS), developed and commercialized by Oxford Nanopore Technologies (ONT), enables the analysis of RNA modifications in native RNA strands [6].

In DRS, native RNA molecules are translocated through individual nanopores embedded in a non-conductive membrane. During library preparation, a helicase is attached to

Received: July 4, 2025. **Revised:** March 12, 2026. **Accepted:** March 23, 2026

each RNA molecule to ensure that migration of the RNA strand through the nanopore occurs at a constant speed. As the RNA strand passes through the pore, it produces characteristic distortions in the electrical current. These electric signals are converted into canonical nucleobases (A, C, G, and U) using machine learning algorithms in a process known as 'basecalling' [3, 7]. Importantly, RNA modifications alter the current intensity signal, thereby providing the basis for their detection in DRS [4, 8, 9]. Over the past years, numerous approaches have been developed to detect a wide range of naturally occurring RNA modifications, including *N*6-methyladenosine (m^6A) [10–18], pseudouridine (Ψ) [18–22], 5-methylcytosine (m^5C) [23], and 2′-O-methylation (Nm) [19, 24, 25], among others.

From a methodological perspective, DRS-based RNA modification detection can be grouped into three broad categories: (i) approaches relying on basecalling 'error' features [10, 11, 14, 22], (ii) methods that employ current-derived features [13, 15, 16, 18, 19, 21, 26], and (iii) modification-aware basecalling models [27, 28]. Basecalling 'error' features exploit the fact that RNA modifications are often incorrectly decoded by standard basecalling models, which were trained to detect the canonical bases (A, C, G, U), resulting in systematic basecalling errors (i.e. mismatches, deletions, and insertions) that become apparent after read mapping. In contrast, current-based methods employ signal-based features, including signal intensity (SI), standard deviation (SD), dwell-time (DT), and/or trace (TR), and apply machine-learning models that have been trained on these features to distinguish unmodified from modified positions. Some of these algorithms operate in a comparison-free mode (*de novo*) without the need for a knockout (KO) sample or an unmodified control. However, recent benchmarks on m^6A detection algorithms indicate that comparison-based methods generally outperform single-mode tools [29, 30]. The third category of methods, modification-aware basecalling models, represents the most recent advancement in the field of modification detection algorithms. These models integrate modification detection directly into the basecalling step and can predict modifications at single-nucleotide and single-molecule resolution. A key advantage of these methods is its independence from aggregated read information, allowing isoform-specific analysis [28] and facilitating detection of RNA modifications even under conditions of low transcript coverage [27].

A central challenge in developing accurate modification-aware models is the selection and preparation of high-quality training data that will be used to train the model. In this regard, early community-driven efforts trained the first modification-aware basecalling models using synthetic sequences with limited sequence contexts [31, 32]. However, these approaches were shown to scale poorly when applied to complex and substoichiometrically modified *in vivo* transcriptomes [27]. Subsequent efforts have used *in vivo* messenger RNA (mRNA) data to train an m^6A-aware basecalling model (the m^6ABasecaller). Because truly modified positions at the per-read level are not known *a priori*, the authors implemented an iterative three-step strategy using gradient boosting to obtain high-confidence per-read labels from *in vivo* data. The obtained labels were then used to train the model, which when applied to *in vivo* human mRNA datasets, was able to predict m^6A sites and their stoichiometry levels with high replicability and low false-positive (FP) rates [27]. More recently, community-driven efforts implemented a novel training strategy, using simulated randomers, to build modification-aware basecallers in a streamlined fashion for different modifications and across sequencing chemistries [33].

In September 2024, ONT released a set of modification-aware basecalling models (m6A_DRACH, inosine_m6A, pseU, and m^5C) that can detect four different RNA modification types, namely for *N*6-methyladenosine (m^6A), inosine (I), 5-methylcytosine (m^5C), and pseudouridine (Ψ), using their proprietary basecalling software 'dorado' (https://github.com/nanoporetech/dorado). These models offer the possibility to predict RNA modifications *de novo*, with single-nucleotide and single-molecule resolution. However, their accuracy, FP rates, and cross-reactivity with other RNA modifications remain unclear. Here, we assess the performance of modification-aware basecalling models released by ONT (see Supplementary Table S1) on synthetic modified and unmodified RNA molecules, as well as on ribosomal RNA from diverse species (Supplementary Table S2). Globally, our results show that these models capture sites at high stoichiometries, but can mispredict a significant number of sites at sub-stoichiometric levels. Moreover, some models show strong cross-reactivity with other RNA modifications found on the same nucleobase, which could lead to inaccurate conclusions if the models are used for *de novo* prediction of RNA modifications. We find that the use of modification-free controls that cover equivalent sequence contexts to the samples of interest [34, 35] are necessary to efficiently remove FPs from nanopore basecalling models. However, we also show that these controls are insufficient to remove all FPs, warranting future improvements to the basecalling models or correction methods. We further demonstrate that neighbouring modifications can confound modification-aware basecalling models. Finally, we show that for those RNA modifications for which modification-aware basecalling models are not available yet, differential base-calling errors and altered current features can be used to identify them [10, 14, 36]. Demonstrating that previously developed approaches for the SQK-RNA002 chemistry can still be used in the context of the new SQK-RNA004 chemistry.

Materials and methods

Escherichia coli culturing and total RNA extraction

Escherichia coli strains used in this work were obtained from the Keio Knockout Collection [37], including the reference wild type strain (BW25113) (Cultek). KO strains have a kanamycin cassette replacing the depleted gene. Strains were plated in LB-agar plates (*E. coli* BW25113) or in LB-agar plates supplemented with 25 μg/ml kanamycin (in the case of *E. coli* KO strains). For each strain, starter cultures were grown in 4 ml LB, incubated at 37°C and 180 rpm. After 6 h, 4 ml cultures were centrifuged in a pre-chilled benchtop centrifuge at 16 000 × *g* 4°C for 5 min. Supernatant was discarded, and pellets were stored at −80 °C overnight. Frozen pellets were thawed on ice, followed by the addition of 400 μl of Trizol (Life Technologies, #15596018). After 5-min incubation at room temperature, 90 μl of chloroform (Vidra Foc, #C2432) was added and mixed thoroughly by inversion. Samples were centrifuged for 15 min at 16 000 × *g* at 4°C. Supernatant was kept and an equal volume of 70% ethanol was added. Afterwards, small RNAs were removed with the

Qiagen RNeasy Mini Kit following manufacturer's recommendations (Qiagen, #74104). Samples were DNAse-treated using Turbo DNAse (Life Technologies, #AM2239) for 15 min at 37°C. Finally, the sample was cleaned with 1X AMPure RNAClean XP beads (Beckman Coulter, A63987). RNA integrity was assessed using TapeStation and quantified using Nanodrop.

Saccharomyces cerevisiae culturing and total RNA extraction

Saccharomyces cerevisiae strains (BY4741) were grown in 4 ml of standard YPD medium (1% yeast extract, 2% Bacto Peptone, and 2% dextrose) overnight at 30°C. The next day, cultures were diluted to 0.0001 OD_{600} in 200 ml of YPD and grown overnight at 30 °C with shaking (250 r.p.m.). When cultures reached the mid-exponential growth phase (OD_{600} 0.5), the cultures were transferred into a pre-chilled 50-ml Falcon tube and centrifuged at 3000 × *g* for 5 min at 4°C, followed by two washes with water, and then pellets were snap-frozen at −80 °C. Snap-frozen yeast pellets were resuspended in 660 µl of TRIzol Reagent (Thermo Fisher Scientific, 15596018) with 340 µl of acid-washed and autoclaved 425–600-µm glass beads (Sigma–Aldrich, G8772). The cells were disrupted by vortexing on top speed for seven cycles of 15 s and chilling the samples on ice for 30 s between cycles. The samples were then incubated at room temperature for 5 min, and 200 µl of chloroform was added. After briefly vortexing the suspension, the samples were incubated for 5 min at room temperature and centrifuged at 14 000 × *g* for 15 min at 4°C. The upper aqueous phase was transferred to a new tube. To precipitate RNA, 1× volume of molecular-grade isopropanol and 1 µl of GlycoBlue co-precipitant (Thermo Fisher Scientific, AM9515) were added and mixed by inverting and incubated for 10 min at room temperature. The samples were centrifuged at 14 000 × *g* for 15 min at 4 °C, and the pellet was then washed with ice-cold 70% ethanol. The pellet was resuspended in 20 µl nuclease-free water after air drying for 5 min on the benchtop, and the RNA purity was measured using a NanoDrop 1000 spectrophotometer. Then, 10 µg of each sample was treated with Turbo DNase (Thermo Fisher Scientific, AM2238) and subsequently cleaned up using a Zymo RNA Clean and Concentrator-5 kit (Zymo Research, R1016) following the manufacturers' instructions. The RNA was eluted in nuclease-free water. RNA concentration was determined using Qubit Fluorometric Quantitation; RNA purity was measured with a NanoDrop 1000 spectrophotometer; and the RNA electropherogram was obtained using Agilent 4200 TapeStation RNA HS ScreenTape Assay.

Mouse models and total RNA extraction

The mice used in this study were of the C57BL/6J strain (Charles River, strain #000664). Animals were maintained on a defined control diet [Special Diets Services, RM1 (P), 801151], housed in standard cages at 22°C–24°C, and given *ad libitum* access to food and water. Lighting conditions were regulated using a 12:12 h light–dark cycle, with lights turned off at 19:00. Adult C57BL/6J mice were euthanized using CO_2. All tissues were promptly dissected, snap-frozen in liquid nitrogen, and stored at −80°C until further use. Animal experiments were conducted in accordance with EU Directive 86/609/EEC and Recommendation 2007/526/EC concerning the protection of animals used for scientific purposes, as implemented by Spanish law 1201/2005. All procedures were approved by the institutional ethics committee.

Human cell line culturing conditions and total RNA extraction

MCF7 (ATCC) cell lines were cultured in Gibco Life Technologies Dulbecco's modified Eagle medium, supplemented with 10% (v/v) fetal bovine serum, and 1% penicillin streptomycin. In order to extract RNA from MCF7 cell cultures, they were grown in 9.6 cm^2 plates at 37°C in a humidified atmosphere containing 5% CO_2 until. Once cells reached 80% confluency, they were first washed with PBS once and then incubated with 1 ml TRIzol at room temperature for 5 min. For each ml of TRIzol, 200 µl of chloroform were added, samples were vortexed for 10 s, incubated for 2–3 min at RT, and spun down at 16 000 × *g* for 15 min at 4°C. The clear top phase was transferred to a new tube, and purified with the miRNeasy kit (Qiagen) according to the manufacturer's instructions. The eluted RNA was quantified by Qubit and RNA integrity was assessed by Agilent Tapestation.

Generation of *in vitro* transcribed rRNAs from total RNA

We generated *in vitro* transcribed rRNA following previously published literature with some minor adjustments [34, 35]. First, total RNA was pre-fragmented to ensure equal coverage across long transcripts by using 150 ng of total RNA, 2µl of RNA Fragmentation Buffer (NEB, E6150S) in a total volume of 20 µl. The reaction was incubated for 1 min at 94°C to obtain a fragment size of ∼500 nt, and stopped by the addition of 2 µl 10× RNA Fragmentation Stop Solution (NEB, E6150S) and transfer to 4°C. Next, 3′ ends were prepared by the addition of 4 µl of FastAP (Thermo Scientific, EF0651), 8 µl of 10× Fast-AP Buffer, in a total volume of 80 µl. The reaction was incubated at 37°C for 15 min followed by a cleanup using the Zymo Clean and Concentrator-5 kit (Zymo Research, R1016). Finally, a short *in vitro* polyadenylation combining the pre-fragmented total RNA with 1 µl of *E. coli* Poly(A) Polymerase (NEB, M0276S), 2µl of ATP (10 mM), and 2 µl of 10× Poly(A) Polymerase Buffer, in a total of 20 µl was carried out. The reaction was incubated at 37°C for 2 min followed by a cleanup using the Zymo Clean and Concentrator-5 kit. For the reverse transcription reaction, 50 ng of input RNA was combined with 1 µl of dNTPs (NEB, N0447S), and 1 µl of the CRTA primer (2 µM, 5′-ACTTGCCTGTCGCTCTATCTTCTTTTTTTTTT-3′) in a total volume of 11 µl, followed by an incubation at 65°C for 5 min before being snap cooled to 4°C. Next, 4 µl of 5× RT Buffer (Thermo Scientific, EP0752), 1 µl of RNAseOUT (Thermo Scientific, 10777019), 1 µl of ddH20, and 2 µl of the SSPII primer (10 µM, 5′-TTTCTGTTGGTGCTGATATTGCTTTVVVVTTVVVVTTVVVVTTVVVVTTTmGmGmG-3′) were added. The reaction was incubated at 42°C for 2 min before 1 µl Maxima H Minus RT (Thermo Scientific, EP0752) was added. The reverse transcription reaction was incubated at 42°C for 90 min followed by heat inactivation at 85°C for 5 min. Next, a polymerase chain reaction (PCR) reaction was performed to enrich for transcripts that underwent successful reverse transcription and template switching. To

this end, 5 µl of complementary DNA (cDNA) from the previous step was added to 25 µl of LongAmp® Taq 2X Master Mix (NEB, M0287S), 2 µl of T7_fwd primer (10 µM, 5′-TAATACGACTCACTATAGCGAGGCGGTTTTCTG TTGGTGCTGATATTGCT-3′), and 2 µl of T7_rev primer (5′-ACTTGCCTGTCGCTCTATCTTC-3′) in a total reaction volume of 50 µl. The PCR consisted of an initial denaturation step at 95°C for 5 min followed by 12 cycles of denaturation (95°C, 15 s), annealing (60°C, 15 s) and extension (65°C, 15 min), with a final extension step at 65°C for 15 min. To remove excess T7_fwd primer, 1 µl of Exonuclease was added and the reaction incubated at 37°C for 15 min, followed by heat inactivation at 80°C for 15 min. Amplified cDNA was cleaned up using 1.8× AMPure XP beads (Beckman Coulter, A63880). Eluted cDNA was used to set up an overnight (16 h) *in vitro* transcription (IVT) using the AmpliScribe T7 High Yield Transcription Kit (Biosearch Technologies, AS3107), following the manufacturer's instructions. IVT libraries were cleaned up using the RNEasy Mini Kit (Qiagen, 74104) following the manufacturer's instructions. To make IVT libraries amenable to nanopore sequencing library preparation, a short *in vitro* polyadenylation was performed using 5 µg of IVT library, 1 µl of *E. coli* Poly(A) Polymerase (NEB, M0276S), 2 µl of ATP (10 mM), and 2 µl of 10× Poly(A) Polymerase Buffer, in a total volume of 20 µl followed by cleanup using the Zymo Clean and Concentrator-5 kit.

Generation of synthetic RNA molecules containing all possible 5-mer sequences

This study used synthetic RNAs known as 'curlcakes', which were designed to include all possible 5-mer sequences while minimizing RNA secondary structures [38], and consist of four *in vitro* transcribed constructs: (i) Curlcake 12 244 bp; (ii) Curlcake 22 459 bp; (iii) Curlcake 32 595 bp, and (iv) Curlcake 42 709 bp. To generate 'curlcake' synthetic RNAs, plasmids (pUC57) were digested with BamHI-HF (NEB, R3136L) and EcoRV-HF (NEB, R3195L) as per manufacturer instructions, followed by a clean-up using phenol/chloroform/isoamyl-alcohol 25/24/1, v/v, pH = 8.05 (Sigma–Aldrich, P3803). Linearized plasmids were used for overnight IVT with the AmpliScribe T7-Flash Transcription Kit (Lucigen, ASF3507). To generate fully modified 'curlcakes', canonical nucleotides were exchanged for the following modified nucleotides: pseudouridine-5′-Triphosphate (pseudo-UTP, Trilink, N-1019) instead of UTP, 5-Methyluridine-5′-Triphosphate (5-meUTP, Trilink, N-1024–1) instead of UTP, *N1*-Methylpseudouridine-5′-Triphosphate (me1Ψ-UTP, Jena Bioscience, NU-890S) instead of UTP, 5-methylcytidine-5′-Triphosphate (5-methyl-CTP Trilink, N-1014), 5-Hydroxymethylcytidine-5′-Triphosphate (5-hme-CTP, Trilink, N-1087), *N*4-Acetylcytidine-5′-Triphosphate (*N4*-Acetyl-CTP, Jena Bioscience, NU-988S) instead of CTP or *N*6-methyladenosine-5′-Triphosphatetriphosphate (*N*6-Methyl-ATP, Trilink, N-1013). To generate 'curlcakes' containing different overall modification levels, mixtures of canonical and modified NTPs were used during IVT (e.g. 1/4 *N*6-Methyl-ATP, 3/4 ATP, to obtain 25% m^6A residues). Following IVT, products were treated with Turbo DNase (Thermo Fisher, AM2238) to remove the template plasmid and cleaned up using the RNeasy MinElute Kit (Qiagen, 74204).

Nanopore direct RNA sequencing of IVTs and total RNA samples

IVT constructs and total RNA samples were polyadenylated using *E. coli* poly(A) polymerase (NEB, M0276S) according to the manufacturer's instructions followed by a clean-up using the RNA Clean & Concentrator-5 kit (Zymo Research, R1013). Concentrations were determined using the Qubit 2.0 Fluorometer. RNA integrity, and polyadenylation success, were determined using an Agilent 4200 TapeStation. Library preparation was carried out according to the manufacturer's instructions (direct-rna-sequencing-sqk-rna004-DRS_9195_v4_revB_20Sep2023-promethion) with minor adjustments to enable highly multiplexed sequencing of multiple barcoded samples [39]. Adapter ligation was carried out for 15 min at room temperature using 100 ng of poly(A)-tailed substrate, 0.5 µl of custom barcode adapter (1.4 µM), 2 µl of NEBNext® Quick Ligation Reaction Buffer (NEB, B6058S), 0.5 µl of SUPERase·In RNAse Inhibitor (Invitrogen, AM2696), and 0.5 µl of concentrated T4 DNA Ligase (NEB, M0202T) in a total volume of 10 µl nuclease-free H_2O. Subsequently, 4 µl of 5× SuperScript IV Buffer, 1µl of dNTPs (NEB, N0447S), 2 µl 0.1 DTT, and 1 µl of SuperScript™ IV Reverse Transcriptase (Thermo Fisher, 18090010) were added and the reaction incubated for 15 min at 50°C before heat inactivation for 5 min at 70°C. Then, RNA:DNA samples designated for sequencing in the same run were pooled accordingly. Separate pools were prepared based on the number of sequencing runs required, and each pool was cleaned up using 1× RNAClean XP beads (Beckman Coulter, A63987) and eluted in nuclease-free H_2O. Twenty-three microlitres of eluate per pool (∼1 µg) was ligated to 6 µl RNA ligation adapter with 3 µl T4 DNA Ligase (NEB, M0202T) and 8 µl of NEBNext® Quick Ligation Reaction Buffer (NEB, B6058S) in a total volume of 40 µl, by incubating at room temperature for 15 min. Libraries were cleaned up using 0.6× RNAClean XP beads (Beckman Coulter, A63987), washed twice with 150 µl of wash buffer (WSB), and eluted in 33 µl of elution buffer (EB). One microlitre of each library was used for quantification with Qubit™ 1× dsDNA High Sensitivity (HS) (Invitrogen, Q33231) and 32 µl were mixed with 100 µl of sequencing buffer (SB) and 68 µl of library solution (LIS). Libraries were run on a primed FLO-PRO004RA flowcell using a PromethION 2 solo sequencer with MinKNOW acquisition software version v24.11.10.

Preprocessing of raw nanopore direct RNA sequencing data

Preprocessing of raw nanopore sequencing data (pod5) was performed using the Master of Pores nextflow pipeline version 4 (MoP4) [40]. First, sequencing data was demultiplexed with SeqTagger (v.1.0c) using the 96-barcode model developed for SQK-RNA004 (-k b96_RNA004) using the default baseQ cut-off of >50 (-b 50) [39]. Next, reads were basecalled with *dorado* (v0.8.0) using each one of the modification-aware basecalling models (m6A_DRACH, inosine_m6A, pseU, and m^5C) with either the hac or sup accuracy settings (v5.1.0). Basecalled reads were aligned to their respective reference with minimap2 (v2.26) [41] using 'sensitive' settings (-y -ax map-ont -k 5) [19] for the highly modified synthetic constructs, and 'default' DRS settings (-y -uf -ax splice -k14) for ribosomal RNA. For parameter choice comparison for mapping,

see Supplementary Fig. S1. Finally, QC plots were generated using NanoPlot (v1.32.1) [42] to visualize read-length versus average-read quality, and custom R scripts (R version 4.3.2) for coverage plots. The b96_RNA004 model is available on GitHub (https://github.com/novoalab/SeqTagger).

Per-read modification probability analysis on synthetic constructs

Per-read and position modification probabilities for each alignment file (BAM) generated by the modification-aware basecalling models were extracted using 'modkit' (v0.4.2, https://github.com/nanoporetech/modkit). For the analysis of the *in vitro* data, 50 000 reads were subsampled from each alignment file (modkit extract full –num-reads 50 000 –mapped-only –reference <ref.fa> <input.bam> <output.tsv>). To generate *in silico* mixes at varying target modification prevalence (0.5–0.001), individual replicates were first pooled and then sampled to the desired ratio using custom R scripts. The resulting tables were further processed to generate modification probability plots, ROC, and PR-ROC plots (Fig. 1). Reads with a minimum baseQ of <10, and reads not matching the reference *k*-mer (query_kmer ≠ ref_kmer) were excluded from the analysis. To perform motif enrichment analysis, we used STREME (v5.5.7) [43] with the entire set of *k*-mers present in curlcakes as the background model and and the underestimated set as test set (modification probability <0.75; Fig. 2B) (streme –verbosity 1 –oc. –rna –totallength 4 000 000 –time 14 400 –minw 5 –maxw 5 –thresh 0.05 –align center).

Per-site stoichiometric analysis of synthetic constructs

To perform per-site analysis, modkit pileup was run with the default filter-threshold as part of the mop_mod MasterOfPores4 workflow (modkit pileup –with-header –filter-percentile 0.1). For the m6A_DRACH model, pileup was run with the additional motif flag (–motif DRACH 2 –ref <ref.fasta>) to ensure that only predictions for the mammalian m^{6}A consensus sequence are considered. Results stemming from different barcodes were collapsed into a single master table using bedtools unionbedg (v2.31). Next, predictions were filtered for the plus strand, a minimum coverage of ≥5, and the central target base, using custom R scripts to perform stoichiometric analysis across a range of modified fractions (Fig. 2C).

Per-site stoichiometric analysis of native and *in vitro* transcribed rRNA

To perform per-site analysis, modkit pileup was run with the default filter threshold (modkit pileup –with-header –filter-percentile 0.1 –motif N 0, where N represents the target base). Finally, *modkit* tables were filtered using custom R scripts to include sites with a minimum valid coverage (N_{val}) of >500 reads. Tables were further annotated with known modified sites. Annotations can be found on the project's github page (https://github.com/novoalab/ONT_basecalling_models/tree/main/ref/annotations). Species silhouettes were obtained from PhyloPics (https://www.phylopic.org/). All scripts have been deposited on the project's github page (https://github.com/novoalab/ONT_basecalling_models/tree/main/).

Alignment and current feature analysis of synthetic constructs and *E. coli* rRNA sequences

To identify RNA modifications in the form of systematic basecalling 'errors', POD5 files from synthetic 'curlcake' datasets and from total *E. coli* RNA samples were processed using the MasterOfPores4 workflow, using either *fast*, *hac*, or *sup* dorado (https://github.com/nanoporetech/dorado) canonical models (v0.8). Modifications were predicted using the EpiNano software [10] to extract alignment features or NanoRMS4 [19] to extract current features (delta signal intensity and dwell time) using the mop_mod module. Differential basecalling errors were converted into *Z*-scores using the mop_consensus module [44].

Results

Expanding direct RNA-sequencing demultiplexing to 96 barcodes for the SQK-RNA004 chemistry

To benchmark ONT modification-aware models across a panel of RNA modifications and varying modification stoichiometries, we generated a total of 192 *in vitro* transcribed constructs, which were pooled into 48 distinct samples (Supplementary Table S2, see also the 'Materials and methods' section). Because sequencing each construct individually would be cost-prohibitive, we sought to implement a multiplexing strategy capable of processing a large number of barcoded samples in a single flowcell. Such an approach would greatly reduce sequencing costs, simplify the experimental workflow, downstream bioinformatic analysis, and minimize potential batch effects. However, no commercial barcoding kit is currently available for the most recent DRS chemistry (SQK-RNA004), and community-driven barcoding efforts [39,45–47] are so far limited to multiplexing up to four barcodes when using the latest RNA chemistry [39, 46].

To address this gap, here we extended previous efforts by training a new model, compatible with the latest SQK-RNA004 kits, to allow demultiplexing up to 96 samples with the current RNA004 chemistry (see Supplementary Table S3 and the 'Materials and methods' section). We found that the newly trained 96-barcoding model (b96_RNA004) reached a precision of 0.993 with 1.0 recall on the validation dataset (Supplementary Fig. S2A). Furthermore, the processing time was not significantly affected compared to our four-barcode model (b04_RNA004), resulting in a mean real-time of 1.3 min per 100 000 reads processed (Supplementary Fig. S2B; see also Supplementary Table S4). Finally, we tested the performance of the b96_RNA004 demultiplexing model on an additional independent test dataset in which barcodes 21, 59, and 88 were absent from the run (see the 'Materials and methods' section). We found that 0.00009% of the total reads were misassigned to either of these barcodes after the default cutoff of baseQ >50 was applied (Supplementary Fig. S2C), suggesting that this model provides a highly accurate and rapid way to sequence up to 96 samples in a single flowcell. Together, these results demonstrate that the b96_RNA004 model provides a robust and scalable solution that enables sequencing large numbers of samples using DRS in a simple and cost-effective manner.

Systematic benchmarking of modification-aware RNA basecalling models with synthetic sequences

We first examined the performance of *dorado* ONT models on datasets with known ground truths and defined per-site modification stoichiometries. More specifically, we employed synthetic RNA constructs, known as 'curlcakes' originally described in [38], which cover all possible 5-mer sequences ($n = 1024$) in multiple copies (median = 10) while minimizing RNA secondary structure. These constructs have been widely applied both to evaluate the detection of diverse RNA modification types on the DRS platform [19, 48], and to train modification-detection algorithms [19, 23, 49, 50].

'Curlcake' constructs were produced with either canonical nucleobases (A, C, G, or U), acting as the unmodified control, or with one modified nucleobase at a time, including those for which modification-aware basecalling models have been trained (m^6A, m^5C, and Ψ), and for four RNA modification types for which there are currently no tailored models available (hm^5C, ac^4C, $m^1\Psi$, and m^5U). Highly multiplexed DRS libraries of synthetic modified and unmodified controls were sequenced in triplicates on a PromethION flowcell (Fig. 1A). The raw sequencing data were then basecalled with either of the four modification-aware basecalling models using the MasterOfPores nextflow workflow (Supplementary Fig. S3A and B) [51], where the model choice was defined by its reference base: (i) m6A_DRACH model: constructs modified with m^6A (at varying concentrations); (ii) inosine_m6A model: constructs modified with m^6A (at varying concentrations); (iii) pseU model: constructs modified with Ψ (at varying concentrations), $m^1\Psi$, or m^5U; and (iv) m^5C model: constructs modified with m^5C, hm^5C or ac^4C. Finally, per-read modification probabilities and per-site modification stoichiometries were extracted using ONT *modkit* software (https://github.com/nanoporetech/modkit) (Fig. 1B) using default parameters (see also the 'Materials and methods' section).

We first evaluated the performance of each model in identifying their designated RNA modification compared to their unmodified counterpart. In the case of m^6A detection, both the 'all-context' model (inosine_m6A) and the 'context-aware' model (m6A_DRACH) were examined. We should note that the latter restricts its predictions to the mammalian m^6A modification consensus motif DRACH (D = A/G/U, R = A/G, and H = A/C/U). We then extracted per-read modification probabilities across all possible 5-mers, finding that the models globally yielded median modification probabilities ranging from 0.83 (m^5C model) to 0.96 (m6A_DRACH model). Thus, performance varied strongly depending on the model, with the m^5C model being the one reporting the overall lowest modification probabilities at the per-read level (Fig. 1C).

To quantitatively assess these differences, we then calculated the performance of each model at per-read level, and built receiver operating characteristic (ROC) curves for each model (Fig. 1D). All models showed accurate predictions globally (ROC-AUC = 0.93–0.97), with the m^5C model demonstrating the lowest area under the curve (AUC) values. Interestingly, we found highly comparable performance of the 'all-context' (inosine_m6A) and the 'context-aware' model (m6A_DRACH) when restricting the sequence context of the former to the DRACH motif, suggesting that the drop in performance observed for the 'all-context' model stems from the larger sequence space it is capable to cover (Fig. 1D).

Modification-aware basecalling models show cross-reactivity against non-target modifications

Current ONT modification-aware basecalling models are trained to identify a given modified base against its canonical counterpart (e.g. m^6A versus A). However, whether these models cross-react with other similar chemical marks that also exist on the same nucleobase remains unexplored. To investigate the potential cross-reactivity of ONT models, we examined their performance on $m^1\Psi$- and m^5U-modified sequences (for the pseU model), and on hm^5C- and ac^4C-modified sequences (for the m^5C model). This analysis revealed substantial cross-reactivity of both models against other modified bases (Fig. 1C and D). Indeed, the pseU model strongly cross-reacted with $m^1\Psi$ modifications (AUC = 0.92), with near-identical AUC values to those obtained against the primary target, pseudouridine (AUC = 0.95), suggesting that this model cannot distinguish the two modification types. Of note, for most *in vivo* analyses, this should still be tolerable, as $m^1\Psi$ is a lowly abundant RNA modification occurring on few RNA biotypes, such as archaeal transfer RNAs [52]. Additionally, this cross-reactivity could be exploited for examining the quality of mRNA vaccine candidates, which are largely generated using $m^1\Psi$ modifications [53], without the need of $m^1\Psi$-specific models. By contrast, m^5U cross-reacted to a lesser extent, yet still significantly better than random (AUC = 0.70). Similar results were observed when examining the cross-reactivity of the m^5C model towards its hydroxylated version hm^5C (AUC = 0.78) and ac^4C (AUC = 0.64) (Fig. 1D). We should note that in the case of m^5C and hm^5C, this cross-reactivity could potentially lead to misinterpretations on *in vivo* data, where the two modifications are known to carry out different functions [54]. Together, these results demonstrate that current modification-aware basecallers exhibit substantial cross-reactivity towards chemically related RNA modifications, even after controlling for neighbouring modified sites.

Evaluation of model performance in biologically representative non-balanced datasets

The prevalence of RNA modifications in benchmarking datasets often does not reflect their true biological occurrence. Benchmarking datasets used to assess model performance are typically constructed using perfectly balanced datasets (i.e. 50% modified reads and 50% unmodified reads). However, the estimated frequency of m^6A, Ψ, and m^5C modifications in mRNAs is estimated to range from 0.1% to 0.6% [55, 56]. Therefore, evaluation of RNA modification detection models should be conducted using non-balanced datasets, which more accurately represent the biological conditions, and would provide a more realistic assessment of the true FP rate that would be observed in biological scenarios.

To this end, we generated *in silico* mixes from our fully modified and unmodified samples at different proportions, obtaining mixes that ranged from perfectly balanced (ratio modified/unmodified = 0.5) to highly unbalanced 'biological-like' scenarios (ratio modified/unmodified = 0.001) (Fig. 1E and Supplementary Fig. S4). Our results showed a consistent decrease in precision and recall with decreased balancing of the dataset across all models. This demonstrates that ONT models show excellent performance in perfectly balanced datasets (Fig. 1D); however, their FP rate dramatically increases when the RNA modification frequency of the

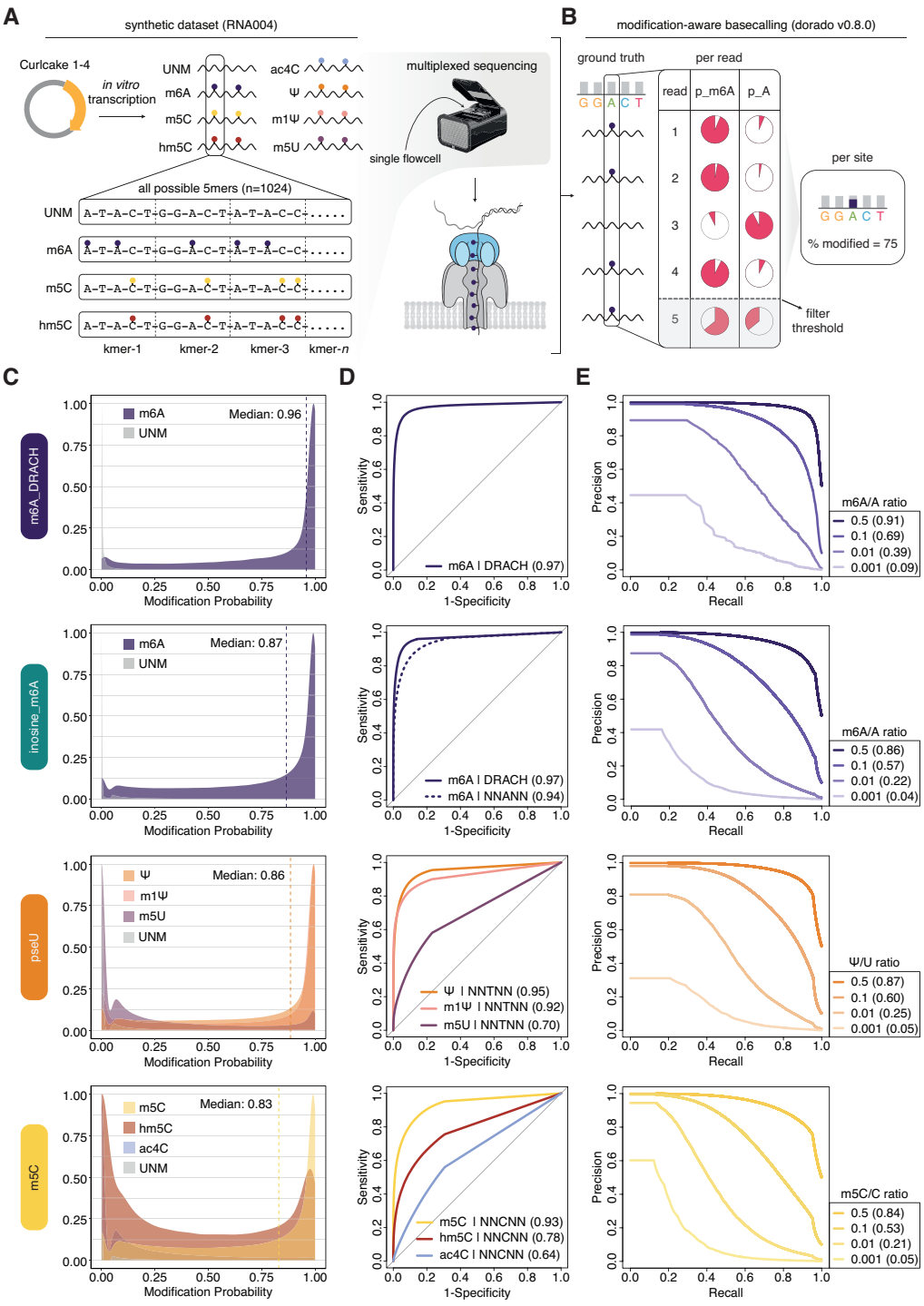


Figure 1. Systematic evaluation of prediction accuracy and cross-reactivity in modification-aware basecalling models using synthetic sequences. (**A**) Schematic representation of the experimental setup used to systematically benchmark modification-aware basecalling models. Synthetic 'curlcake' constructs, which contain all possible 5-mer combinations ($n = 1024$), were generated for a subset of RNA modifications (m^6A, m^5C, hm^5C, ac^4C, Ψ, $m^1Ψ$, and m^5U), including their unmodified counterpart (UNM) and sequenced with the SQK-RNA004 kit in a highly multiplexed fashion using a single flowcell. All constructs were basecalled using *dorado* modification-aware basecalling models (m6A_DRACH, inosine_m6A, pseU, and m^5C). (**B**) Schematic representation of the information provided by modification-aware basecalling models (e.g. single modification m^6A-aware model). First, the model assigns a probability for each read and position of either being modified (e.g. p_m6A) or unmodified (e.g. p_A). Subsequently this information can be collapsed into a per-site prediction using *modkit*, considering a threshold that is calculated per sample (default = 10th percentile). Positions that are below the threshold are excluded from per site analysis (e.g. p_m6A and p_A of read 5 are below the threshold and therefore excluded). (**C**) Per-read modification probabilities reported for each model on fully modified and unmodified sequences. The vertical dashed line represents the median predicted modification probability for the target modification (expected = 100%). (**D**) ROC curves based on per-read modification probabilities for each model. The DRACH motif corresponds to D = A/G/U, R = A/G, and H = A/C/U; N indicates any canonical nucleotide (A/C/G/U). AUC values are provided in brackets. (**E**) Precision–recall (PR) curves based on per-read modification probabilities obtained at different levels of the target modification prevalence (0.5–0.001). PR-AUC values are indicated in brackets. See also Supplementary Fig. S4 for equivalent plots excluding modified bases in the 5-mers, aside from the central position (e.g. BBABB for m^6A, where B = C/G/T).

benchmarking dataset is reflective of that found in *in vivo* datasets (Supplementary Fig. S5; see also Supplementary Table S5). For example, the m6A_DRACH model showed an AUC in the precision-recall curve (AUC-PR) of 0.91 for a balanced dataset (0.5), whereas this model dropped its AUC-PR to 0.09 when using a highly imbalanced dataset (0.001) (Fig. 1E). Notably, even at the very stringent modification probability threshold of 0.99, the proportion of FPs for balanced datasets (ratio = 0.5) was 0.2%–0.3% depending on the model, and this value increased to 63.6%–73.1% when using the most unbalanced design (ratio = 0.001), which in fact resembles the m^6A abundance *in vivo* (Supplementary Fig. S5; see also Supplementary Table S5). Thus, our results demonstrate that highly stringent modification probability thresholding is not sufficient to remove FPs, as under these conditions, the majority of predicted sites in highly unbalanced datasets (ratio = 0.001) represented FPs.

K-mer specific biases of modification-aware basecalling models

Current ONT basecalling models have been trained with synthetic oligonucleotides in which the modified base is surrounded by a degenerate sequence, with the exception of the m6A_DRACH model, which was trained on degenerate oligonucleotides that in the 5-mer region contained the DRACH sequence. For this reason, all models except m6A_DRACH should in principle detect RNA modifications in all sequence contexts. However, whether modifications will be predicted accurately across all sequence contexts has not been systematically benchmarked. Notably, previous works have shown that the alteration of the current signal relative to the unmodified counterpart varies depending on the sequence context [19, 57], suggesting that modifications embedded in certain *k*-mer contexts might be better predicted than others.

To examine the potential *k*-mer preference for detecting RNA modifications of each model, we calculated the median modification probability for each 5-mer containing a single central modified base ($n = 81$), finding that certain *k*-mers were underrepresented, reflected by lower median modification probabilities (Fig. 2A). Next, we performed motif enrichment analysis on those 5-mers whose modification probabilities were lower than 0.75 (see the 'Materials and methods' section), revealing that certain sequence contexts are indeed especially problematic for the detection of modifications. For example, A/U-rich contexts were troublesome for the m^5C model, whereas GC-rich contexts showed decreased modification probabilities for detecting Ψ modifications (Fig. 2B). Troublesome *k*-mers (Fig. 2B, shown in blue) showed modification probabilities profiles that were significantly different from those observed in non-troublesome *k*-mers (Fig 2B, shown in red). This demonstrates that the use of fixed modification probability cut-offs, which is currently becoming a standard in the field [58, 59] will lead to biased representations of modified *k*-mers.

Modification stoichiometry is underestimated by modification-aware basecalling models

Next, we examined the ability of modification-aware basecalling models to make correct predictions in terms of the relative amount of molecules that are modified at a given position, also referred to as percent modified or modification stoichiometry. To examine the accuracy of the models at predicting modification stoichiometry, we built 'curlcake' constructs that contained different proportions of modified NTPs (0%, 12.5%, 25%, 50%, 75%, and 100%), and used the ONT *modkit* tool with default parameters to extract per-site modification levels (see the 'Materials and methods' section).

Our results revealed that all models systematically underestimated the stoichiometries present in the sample, with models predicting a median modification stoichiometry of 72.2%–87.2% for samples that were 100% modified, and with high variance in modification stoichiometry across *k*-mers (Fig. 2C). We should note that the observed stoichiometry values will vary depending on whether the future user is only interested in a subset of *k*-mer sequence contexts. Of note, these results were obtained using default *modkit* parameters, which do not apply fixed per-read modification probability cut-offs in the data. In this regard, the use of fixed per-read modification probability cut-offs (e.g. modification probability >0.95, which are being adopted by the community as a means to remove FPs [59, 60]) will lead to increased underestimation of modification stoichiometries.

Benchmarking of modification-aware basecalling models on *in vivo* total RNA samples

We then evaluated the performance of the models on rRNA sequences from *E. coli* and *S. cerevisiae* for which accurate maps of rRNA modifications exist [61, 62]. For each species, total RNA was extracted and *in vitro* polyadenylated, prepared with standard DRS library preparation, and basecalled using either the pseU or m^5C model (Fig. 3A). Our results showed that the estimates of modification stoichiometry were highly replicable across biological replicates in both species ($R^2 = 0.976$–0.993) (Fig. 3B). However, we should note that a significant proportion of rRNA-modified sites identified were lowly modified (<25%) albeit most sites are expected to be fully modified based on previously published liquid chromatography-tandem mass spectrometry (LC–MS/MS) experiments [62, 63]. This underestimation is also in agreement with our previous observations using synthetic RNAs (Fig. 2C).

Next, we examined whether the sites predicted by the pseU and m^5C models overlapped with annotated Ψ and m^5C sites, respectively (Fig. 3C and D). Indeed, we observed that a significant proportion of predicted Ψ sites (31% in yeast and 21% in *E. coli*) overlapped with annotated Ψ sites; however, several sites containing other modification types (e.g. Um and m^5U) were also captured by the pseU model and thus were mispredicted as Ψ-modified sites (Fig. 3C). In addition, a significant proportion of unmodified sites (66% in yeast and 74% in *E. coli*) were incorrectly predicted as modified. Similarly, the m^5C model captured the two annotated m^5C modified sites in yeast, and two out of three in *E. coli*, but also a vast proportion of FPs, which included both Cm-modified and unmodified sites (Fig. 3D).

In terms of accuracy, most of the high-modification stoichiometry sites were correctly identified by the models (true positives), and these were top-ranked in terms of modification stoichiometry, leading to very high AUC-ROC values (AUC-ROC: 0.920–0.999), but more modest AUC-PR values (AUC-PR: 0.21–0.90) (Supplementary Fig. S6).

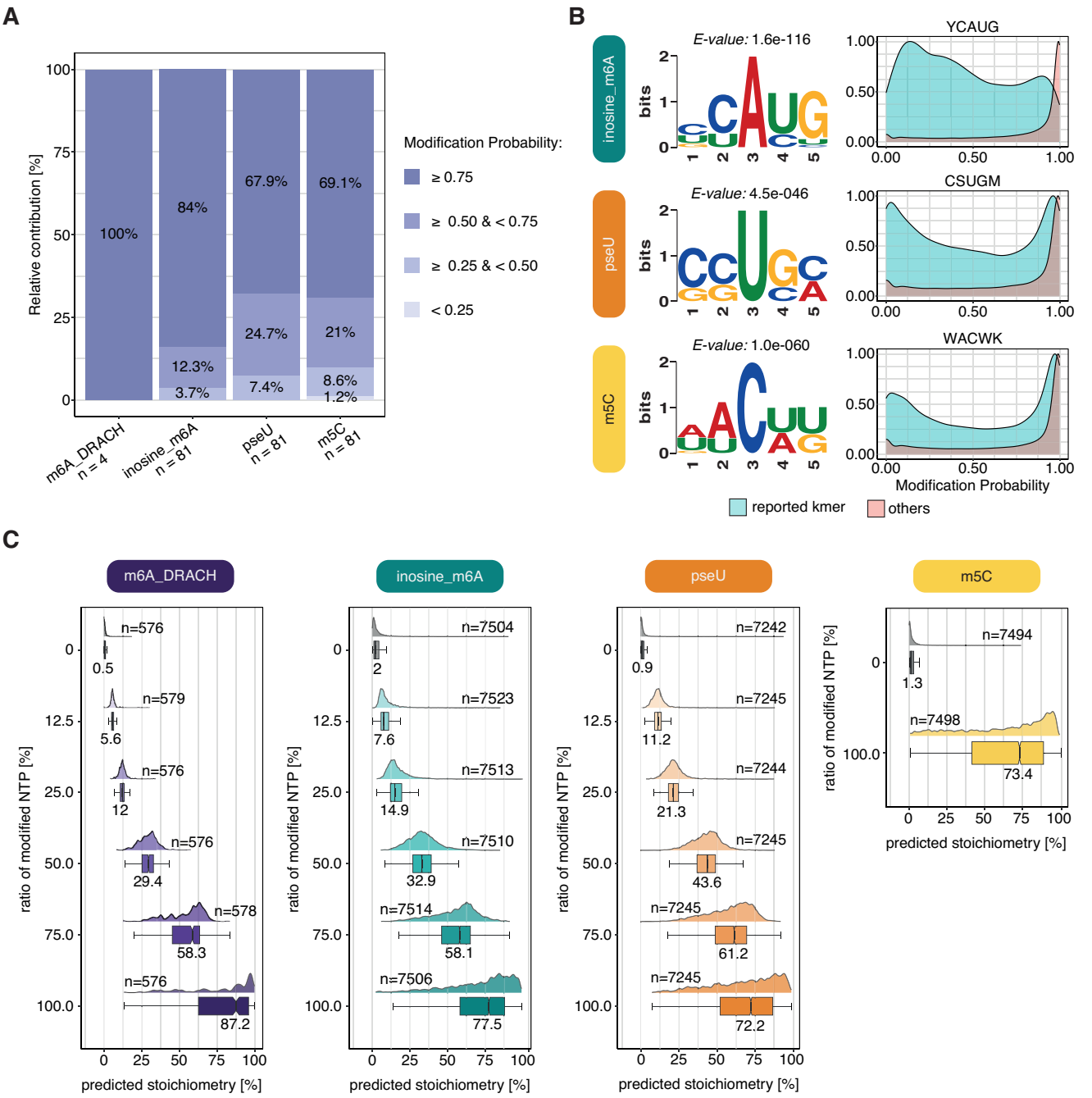


Figure 2. *K*-mer-specific biases of modification-aware basecalling models and stoichiometry estimates. (**A**) Barplot representing the median modification probability of single centrally modified 5-mers for the m6A_DRACH (KGACY, where K = G/T and Y = C/T, $n = 4$), inosine_m6A (BBABB, where B = T/C/G, $n = 81$), pseU (VVTVV, where V = A/C/G, $n = 81$), and m^5C models (DDCDD, where D = A/G/T, $n = 81$). (**B**) Enriched motifs relative to the background distribution of *k*-mers (see the 'Materials and methods' section) and corresponding density plots of modification probabilities for *k*-mers with a median modification probability <0.75. (**C**) Boxplots representing the predicted stoichiometry levels per position across a range of six modification levels (0%, 12.5%, 25.0%, 50.0%, 75.0%, and 100.0% of modified NTP used in each IVT). Median values are indicated below each boxplot with *n* representing the sample size.

Altogether, our results show that *in vivo*, modification-aware basecalling models are able to capture highly modified Ψ- and m^5C-modified sites, but at the expense of a substantial amount of FPs, including other modification types for which the models cross-react with (e.g. Um, m^5U, Cm) as well as unmodified sites that are mispredicted as modified, suggesting that there is a strong need for external fully unmodified controls to remove FP predictions when using modification-aware base-calling models.

Removal of false positives using modification-free synthetic matched controls

Previous works have proposed the use of modification-free control datasets, often generated via IVT, to remove FP predictions in epitranscriptomic studies [64, 65]. We therefore examined whether these constructs would allow flagging mispredicted sites reported by modification-aware basecalling models. To this end, we sequenced both native and IVT control rRNA molecules from *E. coli*, *Mus musculus*, and

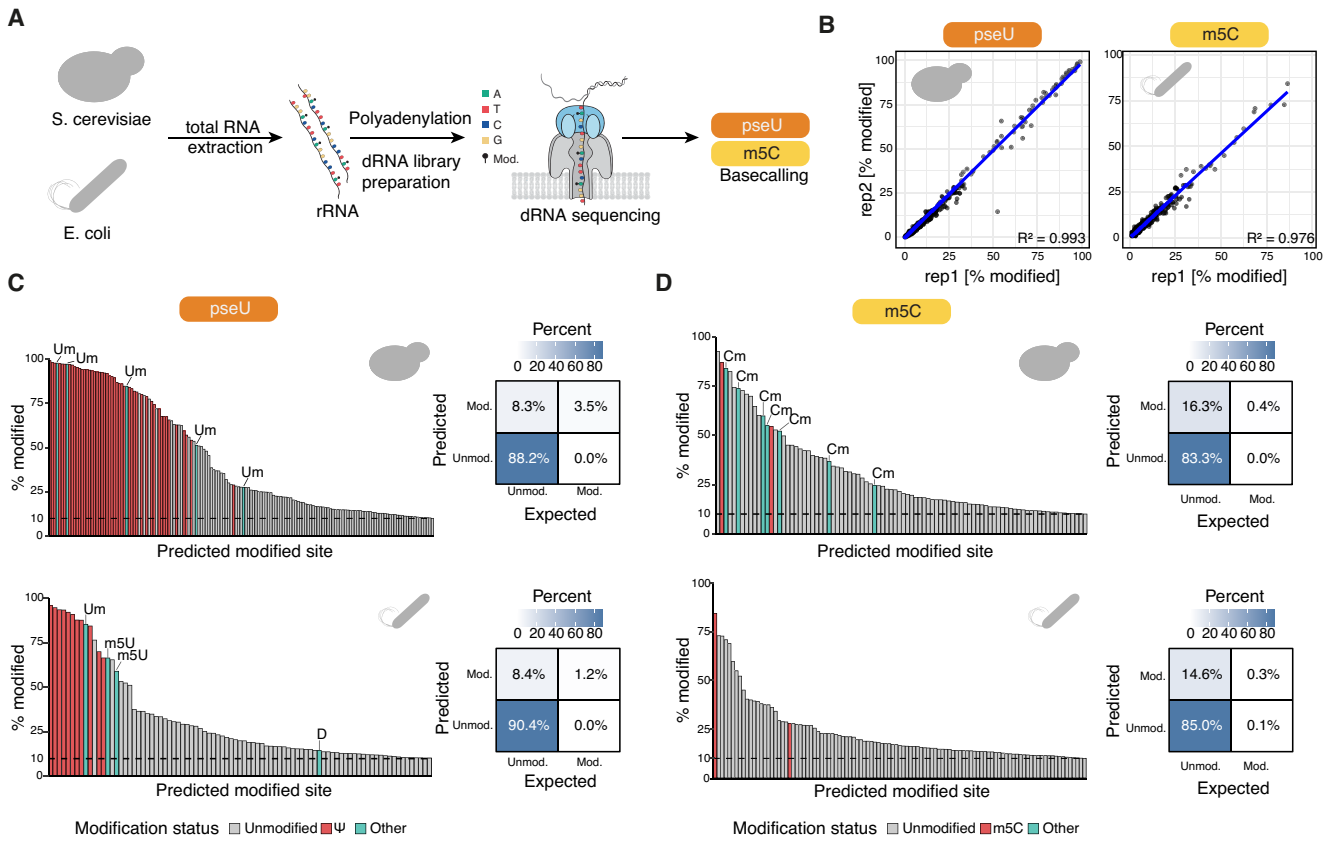


Figure 3. Performance of pseU and m^{5}C modification-aware basecalling models on *S. cerevisiae* and *E. coli* rRNAs. (**A**) Schematic representation of the experimental setup to benchmark modification-aware models in *in vivo* samples. Total RNA from *S. cerevisiae* and *E. coli* was extracted and cleaned, *in vitro* polyadenylated, prepared for DRS using SQK-RNA004 kits. The samples were then basecalled using either N5-methylcytosine (m^{5}C) or pseudouridine (pseU) modification-aware models (see the 'Materials and methods' section) in dorado. Modification information was extracted using the *modkit* software, developed by ONT. (**B**) Scatterplot depicting the predicted modification stoichiometry (% modified) using either the pseU or m^{5}C models, in either *S. cerevisiae* or *E. coli* rRNA, respectively. (**C**, **D**) Histogram depicting the modification stoichiometry for each of the predicted modified sites in rRNAs from *S. cerevisiae* (upper panels) and *E. coli* (lower panels), basecalled with either pseU (**C**) or m^{5}C models (**D**), respectively. Predicted modified sites (by *dorado* + *modkit*) have been coloured depending on whether they are: (i) true-positive sites (red), (ii) non-cognate modified sites (blue), and (iii) unmodified sites predicted as 'modified' (grey). A contingency table matching each histogram, depicting the relative proportion of FPs (mod–unmod), true positives (mod–mod), false negatives (unmod–mod), and true negatives (unmod–unmod), is also shown. For this contingency table, the modification frequency threshold used to identify a site as modified was 10%.

Homo sapiens. We focused on detecting m^{6}A, Ψ, and m^{5}C, which have been reported in rRNAs of all three species [63, 66], and basecalled our samples with the inosine_m6A, pseU, and m^{5}C model comparing predicted modification frequencies in native and unmodified control samples (Fig. 4A, see also the 'Materials and methods' section). Our results revealed that a significant proportion of modifications predicted in native rRNA samples were also predicted in the IVT sequences, confirming that IVT controls can be useful for the removal of FP predictions (Supplementary Tables S6–S8). Notably, replicability of the predictions in the IVT sequences was very high (Pearson's $r = 0.88$–0.99), suggesting that these false predictions are not random but rather systematic, and possibly related to the sequence context itself (Supplementary Fig. S7). Importantly, these findings demonstrate that high reproducibility in predicted modification stoichiometry at a given site does not guarantee that the site represents a true modification.

We then examined whether the use of IVT controls would eliminate all FPs identified in native RNA sequences. To this end, we compared the predicted stoichiometry of native and IVT samples basecalled with each of the three models (Fig. 4B; see also Supplementary Tables S6–S8). We found that, while most of the FP sites were predicted in both native and IVT—i.e. those found on the diagonal (Fig. 4B)—several FP sites were only detected in the native sample. Upon closer examination, we observed that some FP sites exclusively found in native samples corresponded to cross-reactivities against modifications affecting the same nucleobase (e.g. m^{3}Ψ in the case of pseU model) (Fig. 4B and C, and Supplementary Fig. S8), in agreement with our previous observations using synthetic RNA molecules (Fig. 1). Moreover, we identified a third class of FP sites, which corresponded to unmodified residues that had an RNA modification neighbouring them, which we refer to as 'adjacent modification' [e.g. dihydrouridine (D) residues predicted as m^{6}A modifications by the inosine_m6A model, or C$_m$ and m$^6{}_2$A mispredicted as m^{5}C modifications by the m^{5}C model] (Fig. 4B and C, and Supplementary Fig. S8). Altogether, our analysis reveals that there are at least three classes of FP predictions when using modification-aware models: (i) systematic FP sites that also occur in unmodified sequences that can largely be removed using IVT controls; (ii) FP sites caused by cross-reactivity with other RNA modifications (e.g. m^{5}U modifications being called as Ψ modifications); and

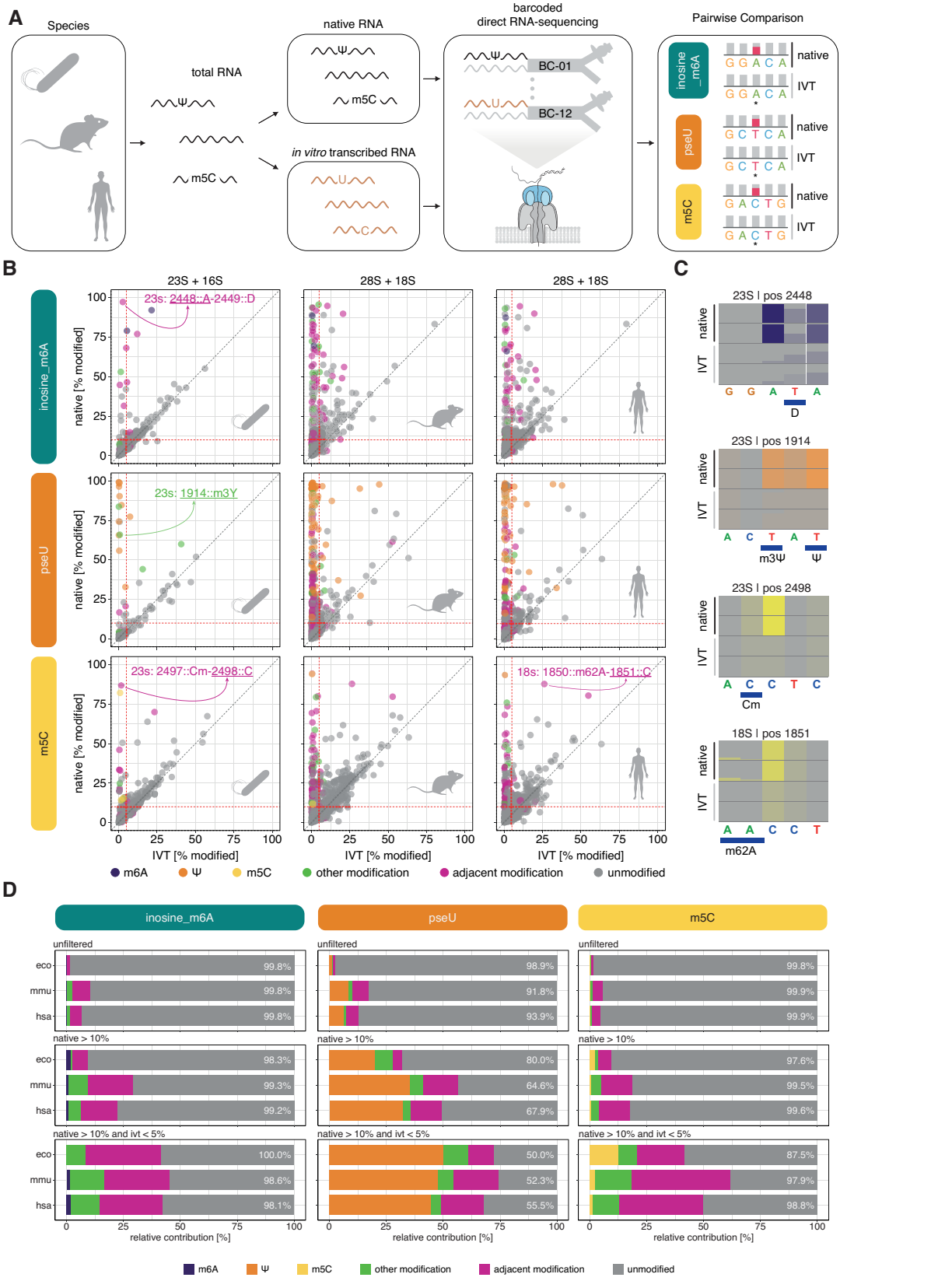


Figure 4. Modification-free controls are essential to remove FPs predicted by modification-aware basecalling models. (**A**) Schematic representation showing the generation of *in vitro* transcribed control samples from the total RNA of three species (*E. coli*, *M. musculus*, and *H. sapiens)*, followed by multiplexed sequencing and pairwise comparisons of the basecalled data using either the inosine_m6A, pseU, or m^5C model. (**B**) Scatterplots comparing the predicted modification stoichiometries in the native and *in vitro* transcribed (IVT) samples for rRNAs found in *E. coli*, *M. musculus*, and *H. sapiens*. Red dotted lines depict the native and IVT thresholds at >10% and <5%, respectively. Individual sites are highlighted to demonstrate the concept of 'other modifications' affecting the same nucleobase (green) and 'adjacent modifications' found ±1 nt of the predicted site (purple). (**C**) IGV

(iii) FP sites caused by neighbouring RNA modifications in other reference bases (e.g. D close to an unmodified A, leading to mispredicting the A as m^6A).

Finally, we quantitatively examined how many FPs would be removed from rRNAs after applying a double threshold, requiring a minimum modification frequency on the native samples (>10%) as well as maximum modification frequency on the IVT samples (<5%). While the use of these thresholds remove a significant proportion of FP predictions in all three tested models, our results also highlight that a large proportion of mispredicted modified sites remained. More specifically, we found that ~55%–58% (inosine_m6A model), ~26%–32% (pseU model), and ~38%–58% (m^5C model) of unmodified sites were still being mispredicted in rRNA molecules, even after excluding sites reported in IVT molecules (Fig. 4D). Notably, in the case of *H. sapiens* we observed a higher median modification stoichiometry for sites with neighbouring modifications than those having a different modified nucleotide from the target one, suggesting that neighbouring modifications, even when found at different bases, can be a major confounder when using modification-aware basecallers, causing the neighbouring unmodified base to be predicted as modified (Supplementary Fig. S8). Altogether, these results show that IVT samples are an essential, yet insufficient, control to remove FP predictions made by modification-aware basecalling models.

Identification of modifications in the absence of modification-aware basecalling models

While modification-aware basecalling models open the possibility to identify modifications in individual molecules, this approach is only applicable to a handful of RNA modifications for which a model is currently available. Therefore, most RNA modifications described to date [66] cannot be identified using modification-aware basecalling models, leaving the majority of RNA modification types unidentified in DRS datasets. To address this, here we examined whether previous approaches, such as the use of 'systematic basecalling errors' [10, 14, 17, 19, 34], and altered current intensity features (e.g. signal intensity and dwell time) [13, 19, 23] can be employed to identify RNA modifications using the SQK-RNA004 chemistry.

Previous works have shown that basecalling 'error' patterns are significantly affected by the choice of canonical model used to basecall the reads [27, 48]. Therefore, we first examined whether the different canonical models released for dorado (*fast*, *hac*, and *sup*) would capture different RNA modification types, using synthetic 'curlcakes' sequenced with the SQK-RNA004 chemistry (Fig. 5A; see also Supplementary Table S9). Our results demonstrate that the *fast* model generates highest global 'error' rates, compared to *hac* and *sup* models at unmodified sites (Fig. 5B). Moreover, all RNA modifications could be detected in the form of basecalling 'errors' across all models. However, these 'errors' were significantly decreased in the case of Ψ, m^5C, and m^6A modifications both in *hac* and *sup* models, with their error rate being close to the background error rate of the unmodified bases (Fig. 5B), suggesting that these modifications have likely been included in the training of 'canonical' models (i.e. the canonical *hac* and *sup* models have learnt to basecall 'm^6A' as 'A'). By contrast, ac^4C, $m^1Ψ$, and m^5U modifications produced high basecalling 'errors' compared to their unmodified counterparts (ΔSumErr), confirming that base-calling 'errors' can be used to detect RNA modifications for which no modification-aware basecalling model exists (Fig. 5C).

Next, we examined whether the 'error' pattern would be centered in the modified site, or in a neighbouring position. To this end, we aggregated the ΔSumErr from all 5-mers, and examined their 'error' patterns along each position of the 5-mer (with 0 corresponding to the modified site) (Fig. 5D). Our results show that, globally, most RNA modifications showed 'error' patterns at the modified site (pos 0), whereas m^5C and hm^5C showed increased errors at the neighbouring position (+1), in agreement with what had been previously observed with SQK-RNA002 chemistry [19, 48]. While the fast model produced the highest 'error' signal (ΔSumErr) for most RNA modification types, we should note that this was not the case for ac^4C modifications (Fig. 5B–D), suggesting that if future users are interested in identifying this modification type, *hac* models will likely better capture their signal.

Finally, we sequenced total RNA from *E. coli* wild type and KO strains lacking specific rRNA modifications at known positions: rsmA KO ($m^{6,6}A$ at 16S:1518 and1519), rsmG KO (m^7G at 16S:527), rsmD KO (m^2G at 16S:966), and rsmE (m^3U at 16S:1498) (see the 'Materials and methods' section) (Fig. 6A; see also Supplementary Table S10). We then examined how these RNA modifications would be detected in the form of 'errors' using by comparing the basecalling 'error' patterns of wild type and KO samples, as well as in the form of altered current features (signal intensity and dwell time) (see the 'Materials and methods' section). Our results showed that all four modification types examined could be robustly identified in the form of basecalling errors (ΔSummedError), alterations in current intensity (ΔSignal Intensity), as well as alterations in dwell time (ΔDwell Time) (Fig. 6B). Thus, we conclude that classical approaches relying on pairwise comparisons remain an effective approach to identify differential RNA modifications by direct comparison of the features from two samples,

snapshots of FP predicted sites highlighted in panel (B), depicting the predicted modification stoichiometries at these sites in two native (dark grey) and two IVT (light grey) replicates. Top: canonical A mispredicted as m^6A due to the presence of dihydrouridine (D) at position 2449 (*E. coli* 23s rRNA). Middle: $m^3Ψ$ at position 1914 (23s) mispredicted as pseudouridine (Ψ). Bottom: canonical C mispredicted as m^5C due to the presence of Cm at position 2497 (23s), and $m^{6,2}A$ at positions 1849–1850 (18s). Tracks were scaled to a maximum coverage of 500 reads. Each position is coloured based on the average modification probability with the height of the bar representing the predicted stoichiometry. (**D**) Barplots depicting the relative fractions of true positives and FPs in rRNA sequences, when using each of the three basecalling models (inosine_m6A, pseU, and m^5C). Predictions given by *modkit* are referred to as 'unfiltered' (top panels); these sites were then filtered to only preserve sites for which the modification frequency was >10% in the native RNA sample (middle panels), and finally sites were further filtered to only keep sites for which the native modification frequency >10% and IVT <5% (bottom panels). Results are shown for rRNAs from *E. coli* (eco), *M. musculus* (mmu), and *H. sapiens* (hsa). True positives are shown as m^6A (purple), pseU (orange), and m^5C (yellow), for each of the basecalling models used, respectively. FPs have been subdivided into: (i) other modifications affecting the same nucleobase (green), e.g. Cm for the m5C model, (ii) unmodified sites adjacent to other RNA modification types (purple), e.g. unmodified C neighbouring an $m^{6,2}A$ site for the m^5C model, and (iii) unmodified bases (shown in grey). The percentages of predicted modified sites that correspond to FP predictions (green + magenta + grey) are indicated within each bar. See also Supplementary Tables S6–S8 for *modkit* predictions on rRNAs using each model (inosine_m6A, pseU, and m^5C), respectively.

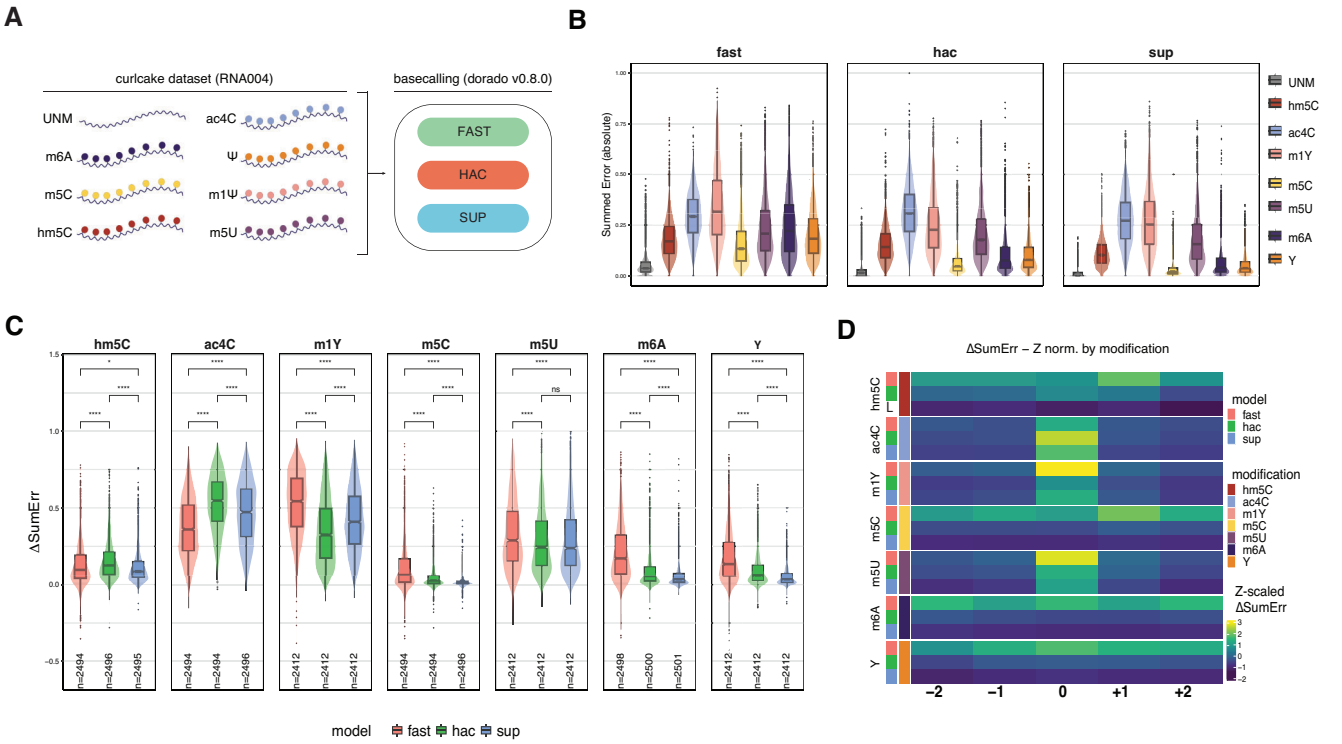


Figure 5. Prediction of RNA modifications using systematic basecalling errors with the SQK-RNA004 chemistry. (**A**) Schematic representation of the 'curlcake' molecules used to benchmark the use of basecalling 'errors' to detect RNA modifications, using either *fast*, *hac*, or *sup* canonical dorado models. (**B**) Summed errors of unmodified and modified bases of curlcake datasets basecalled with either *fast* (left), *hac* (middle), or *sup* (right) dorado canonical basecalling models. (**C**) Comparative analysis of the differential summed error (ΔSumErr) for each RNA modification type examined, calculated by subtracting the 'error' signal obtained at modified and unmodified sites, for each possible 5-mer sequence. (**D**) Analysis of the position at which the 'error' signature is found, for each RNA modification type (Ψ, m^6A, m^5U, m^5C, $m^1\Psi$, ac^4C, and hm^5C) and canonical basecalling model (*fast*, *hac*, *sup*).

in agreement with previous observations obtained using the SQK-RNA002 chemistry [44].

Discussion

The emergence of native RNA sequencing has enabled the possibility to simultaneously examine diverse RNA modifications in full-length reads, with single-nucleotide and single-molecule resolution [27, 67, 68]. This has been greatly advanced by the appearance of modification-aware basecalling models, which allow the detection of RNA modifications during the basecalling step [28]. Following initial attempts by the scientific community [27, 31, 32, 50], in September 2024, ONT released several modification-aware basecalling models for a handful of RNA modification types (m^6A, I, Ψ, and m^5C) (https://github.com/nanoporetech/dorado/releases). However, systematic assessment of their performance, cross-reactivity, and false-discovery rates have only recently begun [60, 69–73].

Here, we systematically evaluated the performance of modification-aware ONT basecalling models using a panel of synthetic constructs covering all possible 5-mers, and containing diverse RNA modification types, including both cognate and non-cognate RNA modifications to the models in question, as well as on *in vivo* and IVT samples from *E. coli*, S. *cerevisiae*, *M. musculus*, and *H. sapiens*. Our results demonstrate that modification-aware base-calling models capture most annotated modified sites (true positives), but mispredict a significant number of sites as modified, leading to 50%–100% FPs (relative to total positives predicted), depending on the basecalling model (Figs 3 and 4D). While we demonstrate that IVT controls can significantly reduce the number of FPs (Fig. 4), a large proportion of incorrectly predicted sites still remain after a double filtering step (Fig. 4). These potential mispredictions could lead to incorrect conclusions when interpreting the biological results, and could further increase controversies in the RNA modifications field, specially in scenarios in which the existence of RNA modifications in certain RNA biotypes has been questioned [65, 74, 75]. Therefore, caution must be taken when interpreting the results to acquire novel biological insights of previously uncharacterized RNA species.

Recent studies have begun to apply highly stringent per-read modification probability thresholds (e.g. mod_prob ≥0.98–0.99) to reduce FP predictions made by modification-aware basecalling models [59, 60]. Although this approach effectively lowers the number of FPs, it does so at the cost of discarding a substantial proportion of true positives. Furthermore, our analysis shows that stringent filtering alone is insufficient to remove FPs in highly imbalanced datasets, such as those typical of *in vivo* conditions (Supplementary Fig. S5), where modified sites occur at much lower frequencies than unmodified sites. Under these non-balanced conditions, we find that the proportion of FP predictions relative to all predictions (FP/P), ranges from 63.6% to 73.1%, depending on the model, even when using highly stringent modification probabilities (mod_prob >0.99) (Supplementary Table S5). These findings highlight the need for caution when using such models for

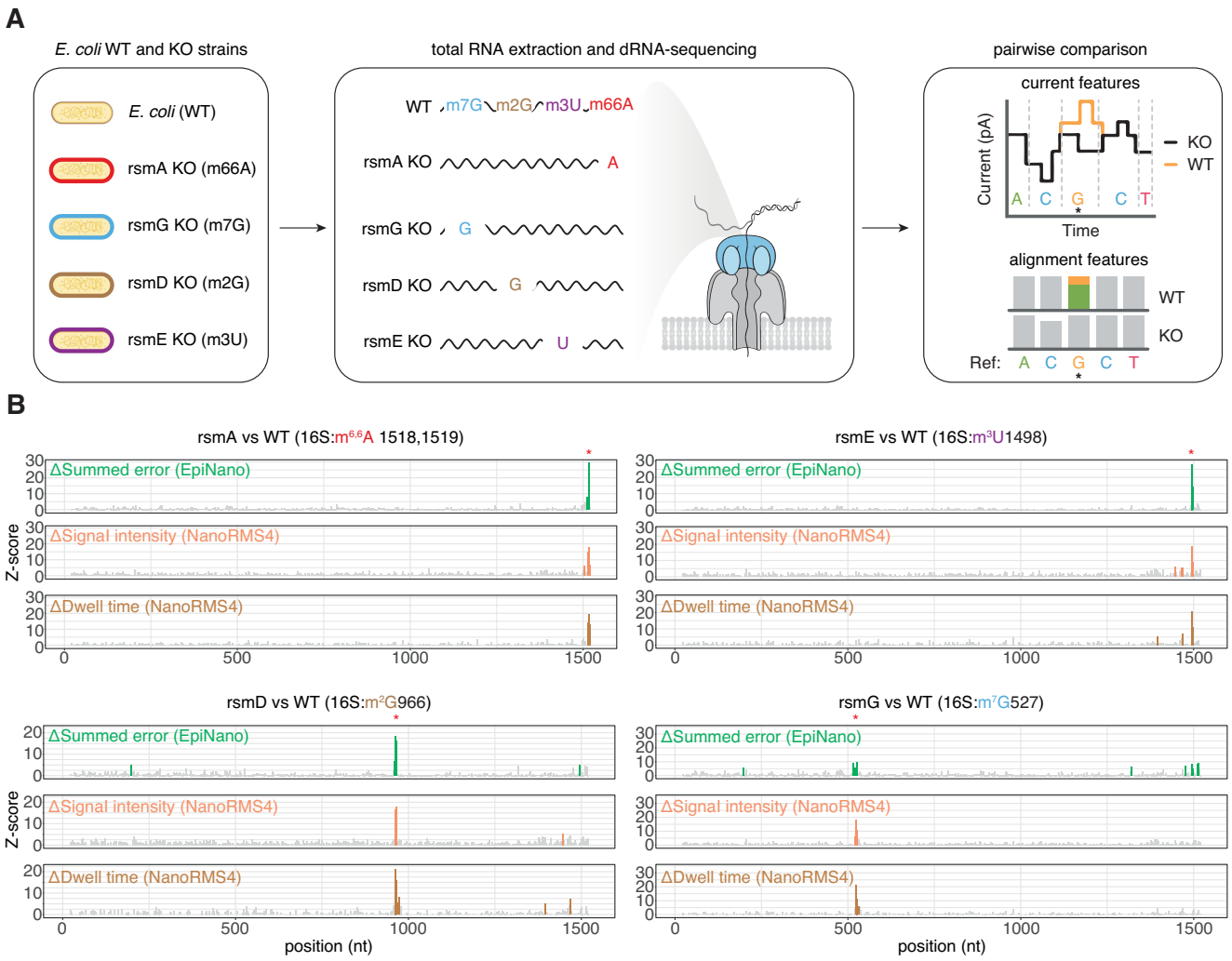


Figure 6. Prediction of RNA modifications using differential basecalling errors and alterations in raw nanopore current features in the SQK-RNA004 chemistry. (**A**) Schematic representation depicting the workflow used to analyse *E. coli* wild type and KO strains. Modifications placed by the indicated enzymes, for which not modification-aware basecallers exist to-date, are shown in brackets. (**B**) Per-position differential modification *Z*-scores along the 16S rRNA transcript, comparing the modification profiles of *E. coli* wild type and each KO strain. For each pairwise comparison, the differential modification profiles were obtained by computing: (i) ΔSummed Error (green), computed using *EpiNano*; (ii) ΔSignal intensity (orange), using *NanoRMS4*, and (iii) ΔDwell Time (brown), using NanoRMS4. The expected differentially modified site for each pairwise comparison is highlighted with a red asterisk.

de novo RNA modification discovery, as stringent probability filtering alone is insufficient to control for FP predictions.

There is a growing consensus about the need to incorporate modification-free transcriptome-wide *in vitro* controls (IVT) as negative controls to improve the accuracy of RNA modification predictions [27, 35, 64]. This is true not only for nanopore-based approaches [34, 65, 76], but also for methods coupled to next-generation-sequencing [64, 77]. Incorporating such controls will likely become routine practice in most laboratories doing RNA modification analyses. We should note, however, that in our study we identify three main sources of FPs predicted by modification-aware basecalling models, some of which cannot be removed using IVT controls: (i) those inherent to the *k*-mer context, which can largely be removed using IVT controls (Fig. 4B); (ii) those caused by cross-reactivity with other RNA modification types that exist at the same reference base, which cannot be removed with IVT controls—but could be potentially removed in the future once modification-aware basecalling models for additional modifications become available—(Figs 1 and 4B and C); and (iii) those caused by cross-reactivity with other RNA modification types present at neighbouring positions, causing misprediction of a modified residue at an unmodified site (e.g. $m^{6,6}A$ neighbouring an unmodified C, leading to m^5C misprediction by the m5C model) (Fig. 4C). Notably, this latter type of FPs will remain even if modification-aware models incorporate additional modifications. This can significantly limit the applicability of modification-aware base-calling models in highly modified RNA biotypes, such as tRNA, rRNA or synthetic RNA constructs.

In this study, we employ 'curlcake' synthetic constructs, which encompass all possible 5-mer sequences, to benchmark basecalling models across all 5-mer sequence contexts. Notably, *hac* and *sup* models trained for the current RNA chemistry (SQK-RNA004) operate with a broader 9-mer signal window. However, recent work has shown that lightweight 5-mer models trained on SQK-RNA004 data achieve similar performance in terms of accuracy and modification

detection relative to their larger 9-mer counterparts, indicating that the majority of informative signal in SQK-RNA004 data remains concentrated within a 5-mer window [78]. Consistent with this observation, previous studies reported minimal differences between read signals produced by newer (9-mer) and earlier (5-mer) RNA chemistries [79]. Taken together, these findings suggest that restricting the analysis to a 5-mer context captures the majority of signal variability required for robust evaluation of basecalling model performance.

Finally, in addition to systematic benchmarking of *dorado* modification-aware models, our study provides a novel demultiplexing model that allows sequencing up to 96 samples in a single flowcell using the latest SQK-RNA004 chemistry. Training such a model was crucial to allow for the cost-efficient sequencing of over a hundred distinct samples (Supplementary Table S2) which were required to systematically benchmark the models across RNA modification types, stoichiometries and RNA biotypes. We should note that ONT recently announced the availability of DRS multiplexing kits in the near future; however, to date commercial multiplexing kits remain unavailable. Community-driven efforts have so far addressed this limitation and have developed accurate algorithms and models compatible with the newest chemistry [39, 46]. Notably, the possibility to multiplex up to 96 barcodes in a single flowcell does not only improve the cost-effectiveness of DRS sequencing, but also facilitates the inclusion of modification-free control samples, such as IVT conditions, which we demonstrate are essential to remove a significant proportion of FP predictions when using modification-aware basecalling models.

Acknowledgements

We thank all the members of the Novoa lab for their insightful discussions.

Author contributions:Gregor Diensthuber (Formal analysis [lead], Investigation [lead], Methodology [lead], Validation [lead], Visualization [lead], Writing—original draft [lead], Writing—review & editing [equal]), Ivan Milenkovic (Data curation [supporting], Formal analysis [supporting], Investigation [supporting], Validation [supporting], Visualization [supporting], Writing—original draft [supporting], Writing—review & editing [supporting]), Laia Llovera (Investigation [supporting], Methodology [supporting], Writing—review & editing [supporting]), Ana Milovanovic (Investigation [supporting], Software [supporting], Visualization [supporting]), Francesco Pelizzari (Investigation [supporting], Software [supporting], Visualization [supporting]), Eva Maria Novoa (Conceptualization [lead], Formal analysis [supporting], Funding acquisition [lead], Investigation [supporting], Project administration [lead],Supervision [lead], Writing—original draft [supporting], Writing—review & editing [supporting]).

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

E.M.N. is a member of the Scientific Advisory Board of IMMAGINA Biotech. G.D., I.M., and E.M.N. have received travel bursaries from ONT to present their work at conferences.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 956810 (GD). E.M.N. and this project received funding from the project PID2021-128193NB-I00 through the MCIN/AEI/10.13039/501100011033/ FEDER, UE; the European Union's Horizon Europe through the European Research Council under the grant agreement numbers 101042103 and 101187456. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. We acknowledge support of the Spanish Ministry of Science and Innovation through the Centro de Excelencia Severo Ochoa (CEX2020-001049-S, MCIN/AEI /10.13039/501100011033), and the Generalitat de Catalunya through the CERCA programme and to the EMBL partnership. We are grateful to the CRG Core Technologies Programme for their support and assistance in this work.

Data availability

Basecalled POD5 files and corresponding aligned BAM files generated in this study have been submitted to the European Nucleotide Archive. This includes data used to train the b96_RNA004 model of SeqTagger (PRJEB90280), and data used for the benchmarking of modification-aware basecalling models (PRJEB91026). A summary of all data sets used in this work can be found in Supplementary Table S2. We would like to note that the triplicate samples for CC1-4 m6A 12.5% were loaded twice on the same flowcell. We hence excluded those samples with less coverage (labelled rep1b-3b) from the analysis. The list of rRNA modifications present in *E. coli*, *S. cerevisiae*, *M. musculus*, and *H. sapiens* rRNAs was obtained from MODOMICS (https://iimcb.genesilico.pl/modomics/sequences/) [66]. Reference sequences for synthetic and *in vivo* samples as well as modification annotations used in this work are available on the project's GitHub repository (https://github.com/novoalab/ONT_basecalling_models/tree/main/ref) and Zenodo (https://doi.org/10.5281/zenodo.19219969). Code to analyse modification-aware basecalled data has been deposited in GitHub (https://github.com/novoalab/ONT_basecalling_models/tree/main).

References

1. Frye M, Jaffrey SR, Pan T *et al.* RNA modifications: what have we learned and where are we headed? *Nat Rev Genet* 2016;17:365–72. https://doi.org/10.1038/nrg.2016.47
2. Suzuki T The expanding world of tRNA modifications and their disease relevance. *Nat Rev Mol Cell Biol* 2021;22:375–92. https://doi.org/10.1038/s41580-021-00342-0
3. Garalde DR, Snell EA, Jachimowicz D *et al.* Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* 2018;15:201–6. https://doi.org/10.1038/nmeth.4577
4. Diensthuber G, Novoa EM Charting the epitranscriptomic landscape across RNA biotypes using native RNA nanopore

sequencing. *Mol Cell* 2025;85:276–89. https://doi.org/10.1016/j.molcel.2024.12.014
5. Fleming AM, Bommisetti P, Xiao S *et al.* Direct nanopore sequencing for the 17 RNA modification types in 36 locations in the *E. coli* ribosome enables monitoring of stress-dependent changes. *ACS Chem Biol* 2023;18:2211–23. https://doi.org/10.1021/acschembio.3c00166
6. Jain M, Olsen HE, Paten B *et al.* The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* 2016;17:239. https://doi.org/10.1186/s13059-016-1103-0
7. Wang Y, Zhao Y, Bollas A *et al.* Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol* 2021;39:1348–65. https://doi.org/10.1038/s41587-021-01108-x
8. Furlan M, Delgado-Tejedor A, Mulroney L *et al.* Computational methods for RNA modification detection from nanopore direct RNA sequencing data. *RNA Biology* 2021;18:31–40. https://doi.org/10.1080/15476286.2021.1978215
9. Wan YK, Hendra C, Pratanwanich PN *et al.* Beyond sequencing: machine learning algorithms extract biology hidden in Nanopore signal data. *Trends Genet* 2022;38:246–57. https://doi.org/10.1016/j.tig.2021.09.001
10. Liu H, Begik O, Novoa EM EpiNano: detection of m6A RNA Modifications Using Oxford Nanopore Direct RNA Sequencing. *Methods Mol Biol* 2021;2298:31–52.
11. Parker MT, Knop K, Sherwood AV *et al.* Nanopore direct RNA sequencing maps the complexity of *Arabidopsis* mRNA processing and m6A modification. *eLife* 2020;9:e49658. https://doi.org/10.7554/eLife.49658
12. Lorenz DA, Sathe S, Einstein JM *et al.* Direct RNA sequencing enables m6A detection in endogenous transcript isoforms at base-specific resolution. *RNA* 2020;26:19–28. https://doi.org/10.1261/rna.072785.119
13. Leger A, Amaral PP, Pandolfini L *et al.* RNA modifications detection by comparative Nanopore direct RNA sequencing. *Nat Commun* 2021;12:7198. https://doi.org/10.1038/s41467-021-27393-3
14. Jenjaroenpun P, Wongsurawat T, Wadley TD *et al.* Decoding the epitranscriptional landscape from native RNA sequences. *Nucleic Acids Res* 2021;49:e7. https://doi.org/10.1093/nar/gkaa620
15. Pratanwanich PN, Yao F, Chen Y *et al.* Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. *Nat Biotechnol* 2021;39:1394–402. https://doi.org/10.1038/s41587-021-00949-w
16. Hendra C, Pratanwanich PN, Wan YK *et al.* Detection of m6A from direct RNA sequencing using a multiple instance learning framework. *Nat Methods* 2022;19:1590–8. https://doi.org/10.1038/s41592-022-01666-1
17. Abebe JS, Price AM, Hayer KE *et al.* DRUMMER—rapid detection of RNA modifications through comparative nanopore sequencing. *Bioinformatics* 2022;38:3113–5. https://doi.org/10.1093/bioinformatics/btac274
18. Chan A, Vries N-d, I.S. S *et al.* Detecting m6A at single-molecular resolution via direct RNA sequencing and realistic training data. *Nat Commun* 2024;15:3323. https://doi.org/10.1038/s41467-024-47661-2
19. Begik O, Lucas MC, Pryszcz LP *et al.* Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nat Biotechnol* 2021;39:1278–91, https://doi.org/10.1038/s41587-021-00915-6
20. Hassan D, Acevedo D, Daulatabad SV *et al.* Penguin: a tool for predicting pseudouridine sites in direct RNA Nanopore sequencing data. *Methods* 2022;203:478–87. https://doi.org/10.1016/j.ymeth.2022.02.005
21. Huang S, Zhang W, Katanski CD *et al.* Interferon inducible pseudouridine modification in human mRNA by quantitative nanopore profiling. *Genome Biol* 2021;22:330. https://doi.org/10.1186/s13059-021-02557-y
22. Piechotta M, Vries N-d, I.S. W *et al.* RNA modification mapping with JACUSA2. *Genome Biol* 2022;23:115. https://doi.org/10.1186/s13059-022-02676-0
23. Acera Mateos P, J Sethi A, Ravindran A *et al.* Prediction of m^6A and m^5C at single-molecule resolution reveals a transcriptome-wide co-occurrence of RNA modifications. *Nat Commun* 2024;15:3899. https://doi.org/10.1038/s41467-024-47953-7
24. Hassan D, Ariyur A, Daulatabad SV *et al.* Nm-Nano: a machine learning framework for transcriptome-wide single-molecule mapping of 2′-O-methylation (Nm) sites in Nanopore direct RNA sequencing datasets. *RNA Biol* 2024;21:560–74, https://doi.org/10.1080/15476286.2024.2352192 .
25. Li Y, Yi Y, Gao X *et al.* 2′-O-methylation at internal sites on mRNA promotes mRNA stability. *Mol Cell* 2024;84:2320–2336.e6, https://doi.org/10.1016/j.molcel.2024.04.011 .
26. Gao Y, Liu X, Wu B *et al.* Quantitative profiling of N6-methyladenosine at single-bfase resolution in stem-differentiating xylem of *Populus trichocarpa* using Nanopore direct RNA sequencing. *Genome Biol* 2021;22:22. https://doi.org/10.1186/s13059-020-02241-7
27. Cruciani S, Delgado-Tejedor A, Pryszcz LP *et al.* De novo basecalling of RNA modifications at single molecule and nucleotide resolution. *Genome Biol* 2025;26:38. https://doi.org/10.1186/s13059-025-03498-6
28. Cruciani S, Novoa EM The new era of single-molecule RNA modification detection through nanopore base-calling models. *Nat Rev Mol Cell Biol* 2025;27:10–8, https://doi.org/10.1038/s41580-025-00896-3
29. Zhong Z-D, Xie Y-Y, Chen H-X *et al.* Systematic comparison of tools used for m6A mapping from nanopore direct RNA sequencing. *Nat Commun* 2023;14:1906. https://doi.org/10.1038/s41467-023-37596-5
30. Maestri S, Furlan M, Mulroney L *et al.* Benchmarking of computational methods for m6A profiling with Nanopore direct RNA sequencing. *Brief Bioinform* 2024;25:bbae001. https://doi.org/10.1093/bib/bbae001
31. Maier KC, Gressel S, Cramer P *et al.* Native molecule sequencing by nano-ID reveals synthesis and stability of RNA isoforms. *Genome Res* 2020;30:1332–44. https://doi.org/10.1101/gr.257857.119
32. Pryszcz LP, Novoa EM ModPhred: an integrative toolkit for the analysis and storage of nanopore sequencing DNA and RNA modification data. *Bioinformatics* 2021;38:257–60. https://doi.org/10.1093/bioinformatics/btab539
33. Fonzino A, Fosso B, Visci G *et al. Ab initio* detection of multiple epitranscriptomic modifications from Oxford nanopore technology direct RNA sequencing data. *Brief Bioinform* 2026;27:bbaf709. https://doi.org/10.1093/bib/bbaf709
34. Tavakoli S, Nabizadeh M, Makhamreh A *et al.* Semi-quantitative detection of pseudouridine modifications and type I/II hypermodifications in human mRNAs using direct long-read sequencing. *Nat Commun* 2023;14:334. https://doi.org/10.1038/s41467-023-35858-w
35. McCormick CA, Akeson S, Tavakoli S *et al.* Multicellular, IVT-derived, unmodified human transcriptome for nanopore-direct RNA analysis. *Gigabyte* 2024;1–13. https://doi.org/10.1101/2023.04.06.535889
36. Naarmann-de Vries IS, Zorbas C, Lemsara A *et al.* Comprehensive identification of diverse ribosomal RNA modifications by targeted nanopore direct RNA sequencing and JACUSA2. *RNA Biology* 2023;20:652–65. https://doi.org/10.1080/15476286.2023.2248752
37. Baba T, Ara T, Hasegawa M *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2006;2:2006.0008. https://doi.org/10.1038/msb4100050

38. Liu H, Begik O, Lucas MC *et al.* Accurate detection of m6A RNA modifications in native RNA sequences. *Nat Commun* 2019;10:1–9.
39. Pryszcz LP, Diensthuber G, Llovera L *et al.* Rapid and accurate demultiplexing of direct RNA nanopore sequencing datasets with SeqTagger. *Genome Res* 2025;35:956–66. https://doi.org/10.1101/gr.279290.124
40. Cozzuto L, Liu H, Pryszcz LP *et al.* MasterOfPores: a workflow for the analysis of Oxford Nanopore direct RNA sequencing datasets. *Front Genet* 2020;11:507358. https://doi.org/10.3389/fgene.2020.00211
41. Li H Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–100. https://doi.org/10.1093/bioinformatics/bty191
42. De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* 2023;39:btad311. https://doi.org/10.1093/bioinformatics/btad311
43. Bailey TL. STREME: accurate and versatile sequence motif discovery. *Bioinformatics* 2021;37:2834–40. https://doi.org/10.1093/bioinformatics/btab203
44. Delgado-Tejedor A, Medina R, Begik O *et al.* Native RNA nanopore sequencing reveals antibiotic-induced loss of rRNA modifications in the A- and P-sites. *Nat Commun* 2024;15:10054. https://doi.org/10.1038/s41467-024-54368-x
45. Smith MA, Ersavas T, Ferguson JM *et al.* Molecular barcoding of native RNAs using nanopore sequencing and deep learning. *Genome Res* 2020; 30:1345–53. https://doi.org/10.1101/gr.260836.120
46. van der Toorn W, Bohn P, Liu-Wei W *et al.* Demultiplexing and barcode-specific adaptive sampling for nanopore direct RNA sequencing. *Nat Commun* 2025;16:3742. https://doi.org/10.1038/s41467-025-59102-9
47. Song J, Lin L-A, Tang C *et al.* DEMINERS enables clinical metagenomics and comparative transcriptomic analysis by increasing throughput and accuracy of nanopore direct RNA sequencing. *Genome Biol* 2025;26:76. https://doi.org/10.1186/s13059-025-03536-3
48. Diensthuber G, Pryszcz LP, Llovera L *et al.* Enhanced detection of RNA modifications and read mapping with high-accuracy nanopore RNA basecalling models. *Genome Res* 2024;34:1865–77. https://doi.org/10.1101/gr.278849.123
49. Vujaklija I, Biđin S, Volarić M *et al.* Detecting a wide range of epitranscriptomic modifications using a nanopore-sequencing-based computational approach with 1D score-clustering. *Nucleic Acids Res* 2025;53:gkae1168. https://doi.org/10.1093/nar/gkae1168
50. Picardi E, Fonzino A, Fosso B *et al. Ab initio* detection of multiple epitranscriptomic modifications from ONT direct RNA sequencing data. *Research Square* 2025; 27:bbaf709. https://doi.org/10.21203/rs.3.rs-6343479/v1
51. Cozzuto L, Delgado-Tejedor A, Hermoso Pulido T *et al.* Nanopore direct RNA sequencing data processing and analysis using MasterOfPores. *Methods Mol Biol* 2023;2624:185–205.
52. Wurm JP, Griese M, Bahr U *et al.* Identification of the enzyme responsible for N1-methylation of pseudouridine 54 in archaeal tRNAs. *RNA* 2012;18:412–20. https://doi.org/10.1261/rna.028498.111
53. Nance KD, Meier JL Modifications in an emergency: the role of N1-methylpseudouridine in COVID-19 vaccines. *ACS Cent Sci* 2021;7:748–56. https://doi.org/10.1021/acscentsci.1c00197
54. Gao Y, Fang J RNA 5-methylcytosine modification and its emerging role as an epitranscriptomic mark. *RNA Biology* 2021;18:117–27. https://doi.org/10.1080/15476286.2021.1950993
55. Liu J, Yue Y, Han D *et al.* A METTL3–METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat Chem Biol* 2013;10:93–5. https://doi.org/10.1038/nchembio.1432
56. Li X, Xiong X, Yi C Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat Methods* 2016;14:23–31. https://doi.org/10.1038/nmeth.4110
57. Fleming AM, Mathewson NJ, Howpay Manage SA *et al.* Nanopore dwell time analysis permits sequencing and conformational assignment of pseudouridine in SARS-CoV-2. *ACS Cent Sci* 2021;7:1707–17. https://doi.org/10.1021/acscentsci.1c00788
58. Hewel C, Wierczeiko A, Miedema J *et al.* Direct RNA sequencing enables improved transcriptome assessment and tracking of RNA modifications for medical applications. *Nucleic Acids Res* 2025;53:gkaf1314. https://doi.org/10.1093/nar/gkaf1314
59. Nicholson LS, Guimarães-Teixeira C, Liu JF *et al.* FTO depletion does not alter m^{6}A stoichiometry in AML mRNA: a reassessment using direct RNA nanopore sequencing. bioRxiv, https://doi.org/10.1101/2025.10.22.681652, 23 October 2025, preprint: not peer reviewed.
60. Hewel C, Wierczeiko A, Miedema J *et al.* Direct RNA sequencing enables improved transcriptome assessment and tracking of RNA modifications for medical applications. *Nucleic Acids Res* 2025;53:gkaf1314. https://doi.org/10.1093/nar/gkaf1314
61. Popova AM, Williamson JR Quantitative analysis of rRNA modifications using stable isotope labeling and mass spectrometry. *J Am Chem Soc* 2014;136:2058–69. https://doi.org/10.1021/ja412084b
62. Taoka M, Nobe Y, Yamaki Y *et al.* The complete chemical structure of Saccharomyces cerevisiae rRNA: partial pseudouridylation of U2345 in 25S rRNA by snoRNA snR9. *Nucleic Acids Res* 2016;44:8951–61. https://doi.org/10.1093/nar/gkw564
63. Taoka M, Nobe Y, Yamaki Y *et al.* Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res* 2018;46:9289–98. https://doi.org/10.1093/nar/gky811
64. Zhang Z, Chen T, Chen H-X *et al.* Systematic calibration of epitranscriptomic maps using a synthetic modification-free RNA library. *Nat Methods* 2021;18:1213–22. https://doi.org/10.1038/s41592-021-01280-7
65. Baquero-Pérez B, Yonchev ID, Delgado-Tejedor A *et al.* N6-methyladenosine modification is not a general trait of viral RNA genomes. *Nat Commun* 2024;15:1964. https://doi.org/10.1038/s41467-024-46278-9
66. Cappannini A, Ray A, Purta E *et al.* MODOMICS: a database of RNA modifications and related information. 2023 update. *Nucleic Acids Res* 2024;52:D239–44. https://doi.org/10.1093/nar/gkad1083
67. Workman RE, Tang AD, Tang PS *et al.* Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat Methods* 2019;16:1297–305. https://doi.org/10.1038/s41592-019-0617-2
68. Huang S, Wylder AC, Pan T Simultaneous nanopore profiling of mRNA m6A and pseudouridine reveals translation coordination. *Nat Biotechnol* 2024;42:1831–5. https://doi.org/10.1038/s41587-024-02135-0
69. Esfahani NG, Stein AJ, Akeson S *et al.* Evaluation of Nanopore direct RNA sequencing updates for modification detection. bioRxiv, https://doi.org/10.1101/2025.05.01.651717, 4 May 2025, preprint: not peer reviewed.
70. Guo Z, Shao Y, Tan L *et al.* Enhanced detection of RNA modifications in Escherichia coli utilizing direct RNA sequencing *Cell Reports Methods* 2025;5:101168. https://doi.org/10.1016/j.crmeth.2025.101168
71. Rübsam FNM, Liu-Wei W, Sun Y *et al.* MoDorado: enhanced detection of tRNA modifications in nanopore sequencing by off-label use of modification callers. *Nucleic Acids Res* 2025;53:gkaf795. https://doi.org/10.1093/nar/gkaf795
72. Zou Y, Ahsan MU, Chan J *et al.* A comparative evaluation of computational models for RNA modification detection using nanopore sequencing with RNA004 chemistry. *Brief. Bioinform.* 2025;4:bbaf404. https://doi.org/10.1093/bib/bbaf404

73. Ohnezeit D, Loliashvili E, Putzel G *et al*. Establishing benchmarks for quantitative mapping of m^{6}A by Nanopore direct RNA sequencing. bioRxiv, https://doi.org/10.1101/2025.11.02.686099, 2 November 2025, preprint: not peer reviewed.
74. Georgeson J, Schwartz S No evidence for ac4C within human mRNA upon data reassessment. *Mol Cell* 2024;84:1601–10.e2. https://doi.org/10.1016/j.molcel.2024.03.017
75. Beiki H, Sturgill D, Arango D *et al*. Detection of ac4C in human mRNA is preserved upon data reassessment. *Mol Cell* 2024;84:1611–25. https://doi.org/10.1016/j.molcel.2024.03.018
76. McCormick CA, Meseonznik M, Qiu Y *et al*. mRNA psi profiling using nanopore DRS reveals cell type-specific pseudouridylation. bioRxiv, https://doi.org/10.1101/2024.05.08.593203, 4 October 2025, preprint: not peer reviewed.
77. Sklias A, Cruciani S, Marchand V *et al*. Comprehensive map of ribosomal 2'-O-methylation and C/D box snoRNAs in *Drosophila melanogaster*. *Nucleic Acids Res* 2024;52:2848–64. https://doi.org/10.1093/nar/gkae139
78. Samarakoon H, Wan YK, Parameswaran S *et al*. Leveraging basecaller's move table to generate a lightweight *k*-mer model for nanopore sequencing analysis. *Bioinformatics* 2025;41:btaf111. https://doi.org/10.1093/bioinformatics/btaf111
79. Kovaka S, Hook PW, Jenike KM *et al*. Uncalled4 improves nanopore DNA and RNA modification detection via fast and accurate signal alignment. *Nat Methods* 2025;22:681–91. https://doi.org/10.1038/s41592-025-02631-4

Received: July 4, 2025. **Revised:** March 12, 2026. **Accepted:** March 23, 2026