Reviewed Preprint
v1 • May 5, 2026
Not revised

Reviewed Preprint
v2 • October 22, 2026
Revised by authors

Calibrated analysis framework for nanopore direct RNA sequencing uncovers cell-specific m⁶A proportions at conserved sites

Denise Ohnezeit[1,*], Elene Loliashvili[2,*], Gregory Putzel[3], Ruth Verstraten[2,4], Adam T Silhavy[1], Jianheng Liu[5], Luke S Nicholson[5], Alejandro Pironti[3], Samie R Jaffrey[5], Daniel P Depledge[1,2,4,6,#], Angus C Wilson[1,#]

[1]Department of Microbiology, New York University School of Medicine, New York, United States • [2]Institute of Virology, Hannover Medical School, Hannover, Germany • [3]Antimicrobial-Resistant Pathogens Program, New York University Grossman School of Medicine, New York, United States • [4]German Center for Infection Research (DZIF), partner site Hannover-Braunschweig, Hannover, Germany • [5]Department of Pharmacology, Weill Medical College, Cornell University, New York, United States • [6]Cluster of Excellence RESIST (EXC 2155), Hannover Medical School, Hannover, Germany

For correspondence: angus.wilson@nyulangone.org

[*] Joint first authors: contributed equally

[#] Joint senior authors: contributed equally, e-mail: *depledge.daniel@mh-hannover.de*; *angus.wilson@nyulangone.org*

Competing interests: No competing interests declared

Funding: See page 46

Reviewing editor: Pedro Batista, National Institutes of Health, National Cancer Institute, United States

© 2026, Ohnezeit et al. This article is distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use and redistribution provided that the original author and source are credited.

eLife Assessment

This **valuable** study compares orthogonal approaches for detecting RNA chemical modifications and provides a helpful framework for improving the reliability of RNA modification identification by direct RNA sequencing. The evidence supporting the technical benchmarking claims is **solid**, with multiple complementary controls, including in vitro-transcribed RNA, chemical inhibition of METTL3, and orthogonal m6A mapping. However, claims regarding relative method accuracy and cell-type- and isoform-specific biological differences remain less well supported due to limited biological replication and validation.

https://doi.org/10.7554/eLife.110672.2.sa3

Abstract

Nanopore direct RNA sequencing (DRS) coupled with Dorado modification-aware basecalling enables mapping of epitranscriptomic modifications including N^6-methyladenosine (m^6A) at the level of individual RNAs. However, the sensitivity, specificity, and reproducibility of this method remain unclear and have only recently begun to be addressed through systematic benchmarking studies. Here, we aimed to establish a best-practice workflow for DRS-based epitranscriptomic analyses. Specifically, we evaluated multiple Dorado versions and models using RNA isolated from primary cells and unmodified *in vitro* transcribed RNAs. We further utilized an m^6A methyltransferase inhibitor as a specificity control. We established that stringent filtering is necessary to reduce false-positive calls and found that Dorado predictions captured an increasing proportion of GLORI sites detected at high m^6A/A proportions. Further, by applying DRS to human primary fibroblasts and HD10.6 neurons, we detected cell type-specific differences in the predicted m^6A/A proportions at conserved sites. Our study thus presents the first systematic comparison of Dorado and GLORI from the same input RNA and expands characterization of the m^6A epitranscriptome to fibroblasts and neurons.

Introduction

Epitranscriptomics describes the occurrence of chemical modifications on RNA that regulate stability, localization, splicing, translation, or decay (Gilbert and Nachtergaele 2023). To date, over 170 RNA modifications are known to occur across all classes of RNA, including eukaryotic

mRNA (Cappannini et al. 2023). Of these, at least thirteen different chemical modifications have been characterized in mRNAs, comprising both cap-proximal and internal modifications (Anreiter et al. 2021). By contrast, current evidence indicates that lncRNAs harbor a more limited repertoire of modifications, primarily N6-methyladenosine (m^6A), 5-methylcytosine (m^5C), and N7-methylguanosine (m^7G) (Yang et al. 2024). One of the best-studied and most abundant internal mRNA modifications is m^6A. Functional studies of m^6A methylases, demethylases, and m^6A binding proteins, as well as sequencing-based mapping methods, have increased our understanding of m^6A as a reversible and dynamic modification that fine-tunes nearly all aspects of mRNA metabolism (Zaccara et al. 2019; He and He 2021; Murakami and Jaffrey 2022; Gilbert and Nachtergaele 2023; Sendinc and Shi 2023).

In addition to m^6A, other widespread and functionally important marks include m^5C, which is less abundant but influences mRNA export and stability, pseudouridine (Ψ), which contributes to mRNA stability, translation and splicing, and adenosine-to-inosine (A to I) editing, which is involved in splicing, stability, and recoding events that contribute to protein diversity (Sun et al. 2023).

Techniques based on short-read (Illumina) sequencing have advanced in recent years to achieve single-nucleotide resolution and quantitative mapping of RNA modifications. The approaches that have emerged to map m^6A can be broadly classified as antibody-based, enzyme-assisted or chemical conversion methods (Moshitch-Moshkovitz et al. 2024).

Early antibody-based techniques (Dominissini et al. 2012; Meyer et al. 2012; Linder et al. 2015) provided the first transcriptome-wide insights into m^6A architecture, albeit at limited resolution. Enzyme-assisted methods (Meyer 2019; Xiao et al. 2023) have improved precision but can introduce enzymatic biases. These biases are avoided with GLORI-Sequencing (GLORI), which selectively deaminates unmethylated adenosines, enabling theoretically unbiased, quantitative, single-nucleotide mapping of m^6A (Liu et al. 2023).

Despite advances, the broad utility of these approaches is limited by a lack of isoform specificity, as well as biases associated with amplification, specificity of antibodies or variable adenosine conversion rates (Diensthuber and Novoa 2025). The short-read length inherent to these approaches also limits the ability to determine whether multiple modifications occur on the same individual RNA molecule. Oxford Nanopore Technologies (ONT) direct RNA sequencing (DRS) offers an alternative to these methods by analyzing native RNA molecules that have retained their modifications (Garalde et al. 2018). This enables modification detection in an isoform-resolved, single-molecule context and, in principle, allows examination of the co-occurrence of multiple modifications along individual RNA molecules. As single RNAs pass through nanopores, signal changes are interpreted to determine canonical and modified residues at single-nucleotide resolution on individual transcripts (Jain et al. 2016; Viehweger et al. 2019). For the now discontinued DRS SQK-RNA002 chemistry used with R9 flowcells, multiple research groups developed algorithms for detecting RNA modifications. These methods can be divided into error rate-based and current signal-based (Abebe et al. 2022). While the recent release of the DRS SQK-RNA004 chemistry and RNA-specific flowcells reduced the utility of many of these tools, introduction of a modification-aware basecaller (Dorado) provides an all-in-one toolkit for basecalling, RNA modification detection, and poly(A) tail length estimation. An early RNA modification-aware Dorado model (v3.0.1) supported m^6A detection exclusively within the DRACH context, while a later model (v5.0.0) introduced detection of Ψ and all-context m^6A. A more recent model (v5.1.0) supports both DRACH-context and all-context m^6A detection, together with Ψ, inosine, and m^5C (*https://github.com/nanoporetech/Dorado*).

In a recent overview of this emerging technology, Cruciani and Novoa highlighted the importance of rigorous evaluation of Dorado's limitations and advocated for the establishment of best practices to ensure consistency and accuracy going forward (Cruciani and Novoa 2026). Furthermore, Diensthuber and Novoa emphasized the need for direct comparisons between Dorado and orthogonal approaches such as GLORI using matched input material, as most benchmarking studies rely on previously published datasets rather than side-by-side experimental comparisons (Diensthuber and Novoa 2025).

To date, several studies have evaluated Dorado in a benchmarking context. Two studies compared Dorado with m6Anet using SQK-RNA004 chemistry and reported improved performance of Dorado, while also incorporating IVT controls to assess false-positive predictions (Zou et al. 2025; Ghobabi Esfahani et al. 2026). Incorporating selective METTL3 inhibition as a complementary, modification-specific negative control could further strengthen the assessment of m^6A detection specificity. Both studies also applied predicted m^6A/A proportion-based filtering to define high-confidence Dorado modification sites, using thresholds of 10% in Zou et al. and 20% in Ghobabi Esfahani et al.; however, the effect of the chosen predicted m^6A/A proportion threshold on site detection and false-positive reduction was not systematically evaluated. Consequently, a comprehensive set of best-practice parameters for Dorado-based modification detection has yet to be established. Notably, although Zou et al. compared Dorado-derived predictions with GLORI-identified sites, the GLORI data were obtained from previously published datasets rather than generated from matched input material, leaving direct comparisons between Dorado and GLORI on the same RNA samples unaddressed (Cruciani et al. 2025; Diensthuber and Novoa 2025; Zou et al. 2025; Ghobabi Esfahani et al. 2026). By directly comparing m^6A detection by Dorado and GLORI on the same RNA inputs, our study overcomes this limitation. Here we systematically benchmark Dorado using poly(A) RNA from primary normal human dermal fibroblasts (NHDFs) and functional human neurons produced through terminal differentiation of HD10.6 cells (Raymon et al. 1999), along with IVT RNA. We included treatment with STM2457, an enzymatic inhibitor of the primary m^6A methyltransferase METTL3 (Yankova et al. 2021), and compared the modification-detection performance of multiple Dorado models released since the tool's introduction. While our analysis is centered on the sensitivity and specificity of m^6A detection in DRS datasets, we also benchmark the ability of Dorado to reliably detect m^5C, Ψ, and inosine. Together, our data provide a robust workflow for DRS-based modification detection that reliably predicts transcriptome-wide, isoform-level, site-specific proportions of RNA modifications. By expanding m^6A mapping on primary human fibroblasts and functional human neurons, we uncover mostly conserved m^6A profiles but reveal increased predicted m^6A/A proportions in neuronal cells.

Results

Systematic benchmarking of ONT's Dorado across human cell types

To evaluate the performance of ONT's basecalling tool Dorado across different human cell types, we selected primary NHDFs and differentiated HD10.6 cells as a neuron-like model. Prior to isolation of total RNA, cells were treated for 48h with a small-molecule inhibitor of METTL3 (STM2457) or with vehicle (DMSO). We ensured that this treatment had no effect on cell viability by performing a CellTiter Glo® assay (Fig. S1). We then isolated poly(A) RNA and subjected the same input material to both DRS and GLORI. The same material was also used to generate IVT RNA (Fig. 1A). Dorado basecalling using software version v.0.9.0 and basecalling model version v.5.1.0 resulted in over 13M reads using poly(A) RNA from NHDFs treated with DMSO or STM2457, while we obtained over 5M and 10M reads from DMSO- and STM2457-treated HD10.6 cells, respectively. The rate of primary alignment to the human genome (hg38) was over 98% for all the sequenced samples. A detailed overview of read statistics generated by DRS can be found in Supplementary Table 1.

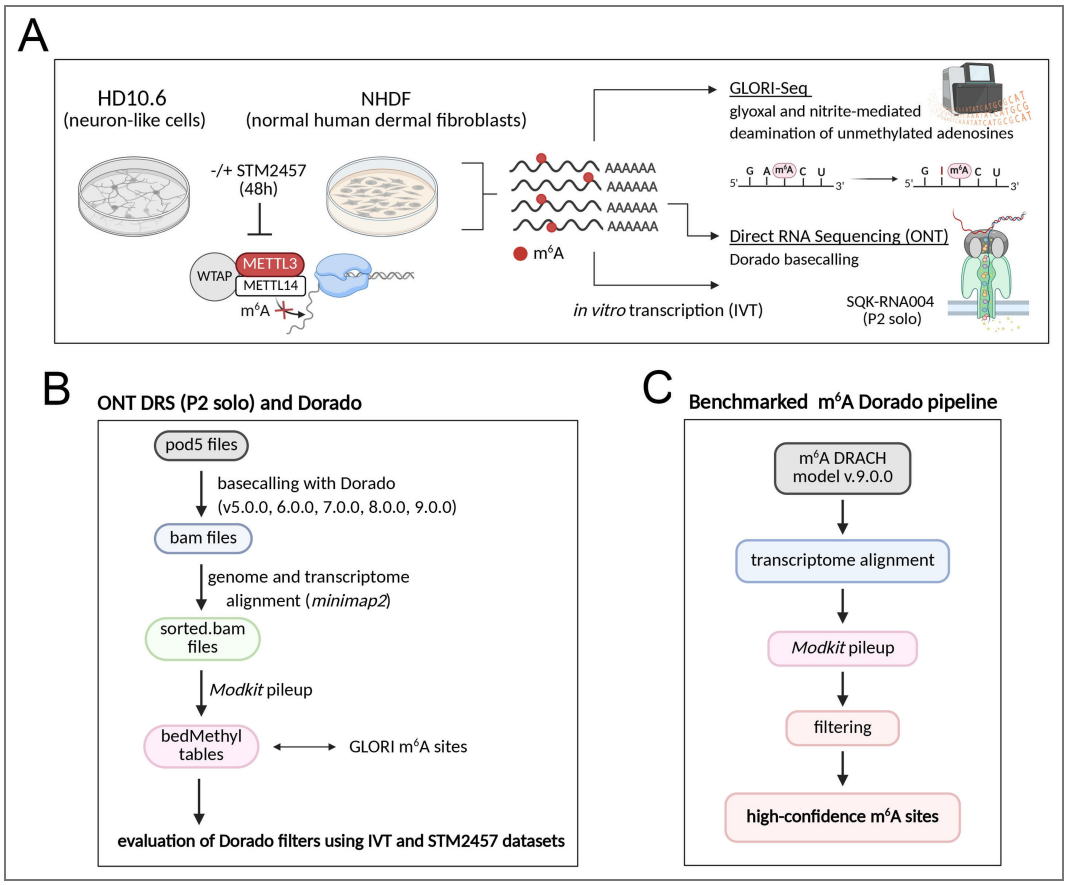


Figure 1. Workflow form A mapping in different human cell types using orthogonal sequencing approaches.

(A) Schematic of the experimental setup. RNA was isolated from NHDFs and differentiated HD10.6 cells treated with STM2457 or DMSO for 48h. Poly(A) RNA was subjected to DRS and GLORI from the same input material. IVT RNA was generated in parallel to assess potential false-positive modification calls. (B) Workflow for benchmarking Dorado basecalling. Reads were processed with Dorado versions v.0.5.0 through v0.9.0, aligned to the human genome or transcriptome, and analyzed with ONT's Modkit tool (Rand 2024). Multiple filtering strategies were tested against IVT controls and METTL3 inhibition (STM2457). Resulting m⁶A sites were compared with GLORI data from the same input RNA. (C) Final benchmarked Dorado pipeline to detect high confidence m^6A sites.

To systematically benchmark Dorado, we performed basecalling using software versions v0.5.0 through v0.9.0. Reads were aligned to the human genome and transcriptome to obtain comprehensive information on isoform-level modification detection. Further, we explored different aspects of detected modifications using the outputs generated by ONT's Modkit tool (Rand 2024). We assessed multiple filtering strategies using base and modification probabilities, as well as predicted modification proportion distributions, and evaluated them based on the IVT data and sensitivity to METTL3 inhibition. By filtering out false-positive calls using multiple control layers, we determined high-confidence m^6A sites (Fig. 1B). Finally, we compared these high-confidence predictions with sites identified by GLORI performed on the same input RNA. Using these results, we developed a benchmarked pipeline that filters m^6A sites predicted by Dorado to minimize false-positive detection without impacting sensitivity (Fig. 1C).

Stringent filtering of modification probabilities removes false positives

During modification-aware basecalling, each nucleotide is assigned a basecall probability score ranging from 0 to 1. Similarly, individual modified nucleotides are assigned a modification probability score in the same range. To establish a baseline for false-positive m^6A predictions, we compared basecall and modification probability distributions between a human genome-aligned IVT poly(A) RNA dataset and DMSO- and STM2457-treated NHDF datasets, all processed using Dorado v0.9.0 (Fig. 2A,B, Fig. S2). Basecall probabilities for adenosine and m^6A modification probabilities from both the all-context and DRACH-context models were grouped into probability bins for comparison.

We observed that 75-90% of A basecalls (in all datasets) have probability scores above 0.8 (Fig. S2A). Since all A bases are considered during all-context m^6A calling, the fraction of m^6A sites is substantially smaller when compared to DRACH-context m^6A sites (Fig. 2A,B, Fig. S2B,C). The modification probability distributions are effectively identical for DMSO-treated, STM2457-treated, and IVT datasets below 0.98, while the profiles differ substantially at probabilities ≥ 0.98. This suggests that sites detected below 0.98 are likely false positives, indicating that only high-probability m^6A calls can be considered reliable. Consistent with this interpretation, METTL3 inhibition with STM2457 markedly reduced the number of detected m^6A sites with modification probabilities ≥ 0.98, compared with DMSO-treated samples, in both the all-context and DRACH-only datasets (Fig. 2A,B, Fig. S2B,C).

To identify specific modification probability cutoffs for each comparison between biological samples and the IVT dataset, we established a statistical framework in which modification probabilities were divided into bins, and enrichment of modification calls in the biological sample relative to IVT was expressed as the log2-transformed sample/IVT ratio. The Pruned Exact Linear Time (PELT) algorithm was then used to identify the point at which this ratio shifted from the baseline distribution, indicating divergence of the biological signal from the IVT background. The modification probability cutoff was subsequently defined as the first bin following the detected change point that exceeded the specified fold-enrichment threshold. This analysis independently supported the use of stringent modification probability cutoffs to distinguish modification-associated signal from the IVT background (Fig. 2A,B).

To assess false-positive modification detection across different modification probability thresholds (0.5, 0.7, 0.9, 0.98, and 0.99), 200,000 reads were randomly subsampled from each dataset in ten iterations, and modification probability distributions were determined for each subsample. False-positive rates were estimated by comparing modification counts in biological samples with those detected in the corresponding IVT controls at each threshold. In NHDF samples, the m^6A DRACH sites showed a relatively low false-positive rate (~10%) even at a probability threshold of 0.5, which decreased further at more stringent thresholds (Fig. S3A). By contrast, the m^6A all-context model showed substantially higher false-positive rates (~35% at ≥0.5), decreasing to <5% at the highest thresholds. Analysis of discrete modification probability ranges showed a similar trend (Fig. S3B), with false-positive rates decreasing significantly at higher modification probabilities. Statistical significance was assessed using pairwise comparisons derived from linear mixed-effects

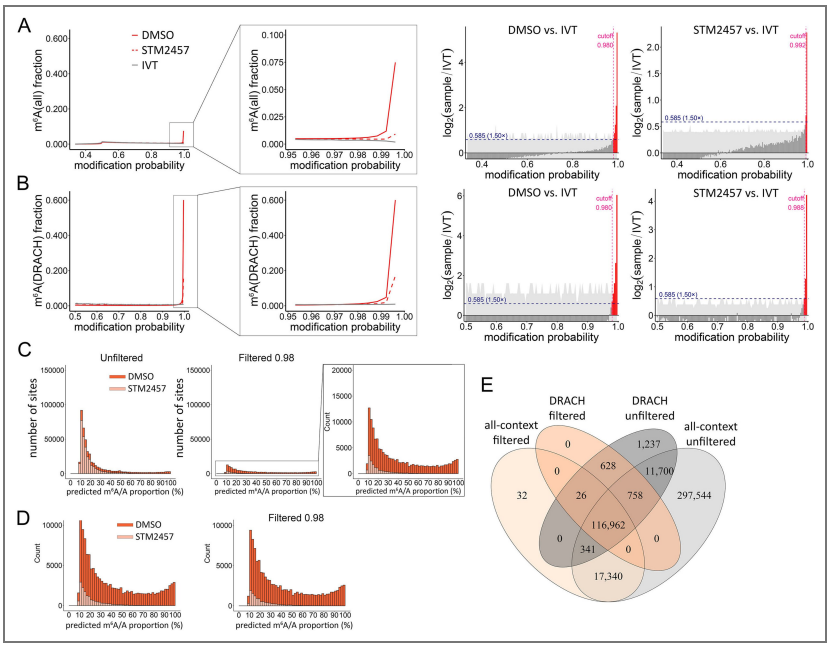


Figure 2. Modification probability of 0.98 and predicted m A/A proportion score of ≥ 10% are necessary cutoffs to identify m A sites accurately.

(A) (left) Dorado v0.9.0-generated modification probability distributions for all-context m^6A from genome-aligned DMSO- and STM2457-treated NHDFs plotted against the distributions obtained from the IVT NHDF poly(A) RNA. DMSO probability score distribution is shown as a red solid line, STM2457 – red dashed line, IVT – grey solid line. (right) Bar plots showing the per-bin log2(sample/IVT) ratio of modification probability fractions for the DMSO- and STM2457-treated NHDFs conditions relative to the unmodified IVT reference. The horizontal navy dashed line marks the fold-enrichment threshold used to call enrichment. Bars are coloured by significance relative to the fold-enrichment threshold (red, log2 ≥ 1.5-fold; dark grey, not significant). The light grey ribbon indicates the 95% bootstrap confidence interval around the log2 ratio estimate at each bin. The vertical pink dashed line indicates the derived modification probability cutoff above which sites are considered genuinely modified. (B) Same as in (A), but for DRACH-context m^6A sites. (C) Predicted m^6A/A proportion (≥ 10%) distributions for all-context m^6A from DMSO-(in red) or STM2457-treated (in pink) NHDF datasets (genome-aligned) are plotted using a 0.98 modification probability cutoff and a filter for ≥ 20 reads. (D) Same as in (C), but for DRACH-context m^6A sites. (E) Overlap of m^6A sites between all-context and DRACH models (A basecall accuracy ≥ 80%, coverage ≥ 20 reads, predicted m^6A/A proportion ≥ 10%), comparing non-filtered to filtered (modification probability > 0.98) outputs.

models, with sample included as a random effect and Benjamini-Hochberg correction for multiple comparisons. The reduction in false-positive rates of all-context m^6A detection was less pronounced following STM2457 treatment, indicating that false-positive signal becomes proportionally more prominent as the number of genuine modifications decreases. Similar patterns were observed in HD10.6 cells (Fig. S4A,B).

Furthermore, we examined modification probability distributions for Ψ, inosine and m^5C, each of which can be basecalled with Dorado v0.9.0 (Fig. S5). For Ψ and I, the modification probability distributions for the STM2457 and DMSO datasets only diverge from the IVT dataset at ≥ 0.99 modification probability, supporting the use of this stringent threshold to reduce false-positive detection (Fig. S5A,B). Consistent with this, estimated false-positive rates decreased with increasing modification probability, although they remained comparatively high (~20-30%) even at stringent thresholds (Fig. S5A,B, Fig. S6A, Fig. S7A). Analysis across discrete modification probability intervals showed the same trend, with progressively lower error rates at higher probabilities (Fig. S5A,B, Fig. S6B, Fig. S7B). By contrast, m^5C modification probability distributions were similar between all three datasets, and no filter could be adequately established (Fig. S5C). While this likely reflects a low rate of m^5C installation, it also indicates that the m^5C model utilized by Dorado requires further refinement.

Modification probability distributions have similar profiles for transcriptome-aligned NHDF data (Fig. S8) as well as genome- and transcriptome-aligned data from HD10.6 cells (Fig. S9, Fig. S10). Of note, the fraction of high-probability inosine sites is increased in STM2457-treated cells, compared to DMSO-treated ones in genome-aligned data (Fig. S5B, Fig. S9D), which points to increased numbers of retained introns in the METTL3-inhibited sample. Accordingly, this effect is absent in transcriptome-aligned data (Fig. S8D, Fig. S10D).

While IVT transcriptomes serve as excellent negative controls, we also considered whether the RNA calibration strand (RCS) that is added during DRS library preparation could serve a similar function. The RCS is a synthetic Enolase II RNA derived by IVT that, similarly to the IVT transcriptomes, is devoid of RNA modifications. Despite the reduced diversity of sequence in the RCS data, the overall results were comparable to those from the IVT transcriptome data (Fig. S11A,B, Fig. 2A,B), demonstrating that the RCS can be used as a "built-in" negative control for modification detection with DRS. Finally, we determined that basecall and modification probability distributions remain generally consistent between different Dorado models, although more m^6A and Ψ sites are detected with high modification probability in Dorado model v5.1.0 compared to v5.0.0 (Fig. S11C-E).

Based on our observations, we propose a base probability cutoff of 0.8 and a modification probability cutoff of 0.98 for m^6A (and 0.99 for Ψ and inosine) to reduce false-positive calls.

Filtering by predicted m A/A proportion further eliminates potential false positives

We next examined the predicted m^6A/A proportion distributions for all m^6A sites in both modification probability-unfiltered and filtered NHDF datasets (Fig. S12). In addition to applying base and modification probability cutoffs, we restricted analyses to sites corresponding to adenosine residues in the reference sequence used for alignment, an approach also reported by Esfahani and colleagues (Ghohabi Esfahani et al. 2026) thereby further reducing the likelihood of false-positive calls. We additionally required a minimum valid coverage of 20 reads per site, which is also consistent with the criteria applied in the same study. Valid coverage was defined as the sum of reads supporting the canonical adenosine base and reads supporting an m^6A modification call (with ≥0.98 modification probability score). Unfiltered datasets showed that Dorado identifies 11.5x more m^6A sites with the all-context model than with the DRACH model (Supplementary Table 2), consistent with the prior study by Cruciani and colleagues (Cruciani et.al. 2026), reporting 20x more all-context m^6A sites than DRACH-context. Furthermore, we showed

that the vast majority of m^6A sites, 92% for all-context and 75% for DRACH-context, have predicted m^6A/A proportion values <10%, with less than 1% of sites showing high (>50%) predicted proportions.

A similar analysis of the IVT transcriptome revealed large numbers of false-positive m^6A sites with predicted m^6A/A proportion values below 10%, most of which were eliminated by application of our filters (Fig. S13A,B). However, the remaining presence of low-proportion false-positive sites indicated that an additional layer of filtering to exclude sites with <10% proportion is beneficial. Notably, this filter was also adopted in the benchmarking study by Zou and colleagues (Zou et al. 2025). An additional argument supporting our filtering strategy is the observation that the effect of treatment with the METTL3-specific inhibitor STM2457 is much more pronounced at m^6A sites with ≥10% predicted m^6A/A proportion (Fig. 2C,D; Supplementary Table 2), regardless of the motif context. Applying modification probability and per-site proportion cutoffs substantially reduced the difference between the numbers of m^6A sites detected with all-context and DRACH-context modes of Dorado (Supplementary Table 2). Application of our modification probability filtering strategy removed over 70% of the sites (Fig. 2E), regardless of the context, almost all of which showed predicted m^6A/A proportions of <10%, without significantly changing the numbers of high-proportion sites detected. Similar results were also observed for the transcriptome-aligned NHDF and HD10.6 distributions (Fig. S14, Fig. S15; Supplementary Table 2).

To determine whether our filtering strategy could be applied to other RNA modifications detected by Dorado, we examined the distributions of predicted Ψ/U and inosine/A proportions in human genome-aligned DMSO- and STM2457-treated NHDF and HD10.6 datasets, before and after applying the ≥0.99 modification probability threshold. Similar to m^6A (Fig. S12), the majority of detected Ψ and inosine sites exhibited low predicted modification proportions, with 92% of Ψ sites and 96% of inosine sites showing proportions below 10% (Fig. S16 for NHDFs, Fig. S17 for HD10.6 cells). This predominance of low-proportion Ψ sites is consistent with a recent study using 2-bromoacrylamide-assisted cyclization sequencing (BACS), which reported that 82% of Ψ modifications in human mRNAs occurred at proportions below 20% (Xu et al. 2024).

Importantly, analysis of the corresponding IVT datasets further supported the use of the ≥0.99 modification probability threshold for Ψ and inosine (Fig. S13C,D). Applying this threshold substantially reduced the number of sites detected in the IVT controls compared with the unfiltered data. For inosine, false-positive predictions were present in the unfiltered IVT data even at high (>50%) predicted modification proportions and were markedly reduced following modification probability filtering (Fig. S13D). These findings indicate that stringent modification probability filtering is necessary to reduce background predictions for Ψ and inosine before interpreting their estimated per-site proportions. We have thus demonstrated that setting thresholds for basecall modification probabilities provides a robust method for identifying and filtering false-positive modification calls.

Ψ and inosine sites exhibit the highest degree of modification-specific basecalling mismatches

An important consideration when interpreting the predicted modification proportions is that these estimates do not account for all reads overlapping a given position. Modification proportions are calculated from reads in which the corresponding reference base is basecalled as the expected canonical nucleotide and is subsequently assigned as either canonical or modified (above the selected modification probability threshold). Reads in which the position is basecalled as a different nucleotide, contains a deletion, is assigned another modification, or does not meet the selected modification probability threshold are therefore excluded from the proportion calculation. For the DRACH-context m^6A model, modification calling additionally depends on preservation of the required sequence context, such that reads containing a non-matching base within the motif may also be excluded. Consequently, modification-associated effects on canonical basecalling could introduce bias into the reported modification proportions.

To assess the extent of basecalling mismatches, we examined alternative nucleotide calls at identified modification sites in the filtered DMSO-treated NHDF and HD10.6 datasets and compared, for each site, the percentage of reads containing a non-reference basecall with the predicted modification proportion (Fig. S18). For most sites, non-reference basecalls represented <25% of overlapping reads. Sites at which ≥25% of reads contained a non-reference basecall were comparatively uncommon and were predominantly associated with low predicted modification proportions (<10%). We examined the composition of alternative nucleotide calls at these sites across the different modification types. Notably, in both NHDF and HD10.6 cells, C accounted for the majority of alternative basecalls at Ψ sites, whereas G predominated at inosine sites (Fig. S18).

These observations indicate that basecalling errors can exclude a subset of reads from modification proportion calculations, with the extent and nature of this effect differing between modification types. The predicted modification proportions may therefore underestimate the true proportion of modified molecules at reported sites, as reads containing a modification but basecalled as an alternative nucleotide would not contribute to the modification estimate. This dependency may also result in some modification sites being missed altogether when basecalling errors reduce the number of reads eligible for downstream modification calling. Thus, the reported proportions should be interpreted as only estimates derived from reads satisfying the respective basecalling and modification-calling criteria.

RNA modification detection rates differ between Dorado models

Dorado basecalling relies on models generated using Remora (*https://github.com/nanoporetech/remora*) that are frequently updated. This raises an important question as to whether these updates produce notably different outputs that could impact biological interpretation. To address this, we utilized multiple Dorado versions and Remora models to perform modification calling of our NHDF dataset. We restricted our analyses to m^6A and Ψ as these are the only modifications that could be detected in previous models. Here, we observe that predicted m^6A and Ψ proportion distributions are generally consistent between different Dorado versions and models (Fig. S19), albeit with greater numbers of low predicted proportion (< 10%) sites detected in older releases. There is a large overlap of the detected ≥ 10%-proportion of m^6A (DRACH and all-context) and Ψ sites between versions/models, although there are a lot of unique all-context m^6A and Ψ sites identified with just model v5.0.0 and not v5.1.0, that are reduced after applying our modification probability filter (Fig. S20A,B,C). Ψ sites uniquely detected with different models are found in slightly different sequence contexts (Fig. S20C,D). Thus, while RNA modification detection with Dorado is sensitive to changes in the underlying models/versions, this can be effectively controlled through the application of filters that only retain sites with high-confidence modification probability calls.

GLORI exhibits reduced specificity but converges with Dorado at sites with high predicted modification proportions

To assess the agreement of DRS Dorado-derived predictions with an orthogonal m^6A detection method, we conducted GLORI on the same NHDF RNA preparations. To ensure a fair comparison of m^6A sites across the two methods, we applied the following filtering strategies: For Dorado (DRACH model), this included thresholds on modification probability (A ≥ 0.8, m^6A ≥ 0.98) and read coverage (≥ 20). For GLORI, we used filters based on coverage (≥ 20) and statistical significance (p-value ≤ 0.05). Next, we extracted m^6A sites that were present in both DMSO and STM2457-treated samples and removed all sites with ≤ 10% predicted proportion in DMSO (Fig. 3A). This stepwise strategy reduced false positives and enabled a fair comparison of the high-confidence m^6A sites identified by both approaches.

As shown in the violin plots, Dorado identified similar numbers of m^6A sites using all-context and DRACH-context modification calling (81,012 and 72,372 sites, respectively), with predicted modification proportions strongly reduced upon STM2457 treatment in both cases (Fig. 3B). The two Dorado approaches differ in the sequence context in which m^6A modifications are called: all-

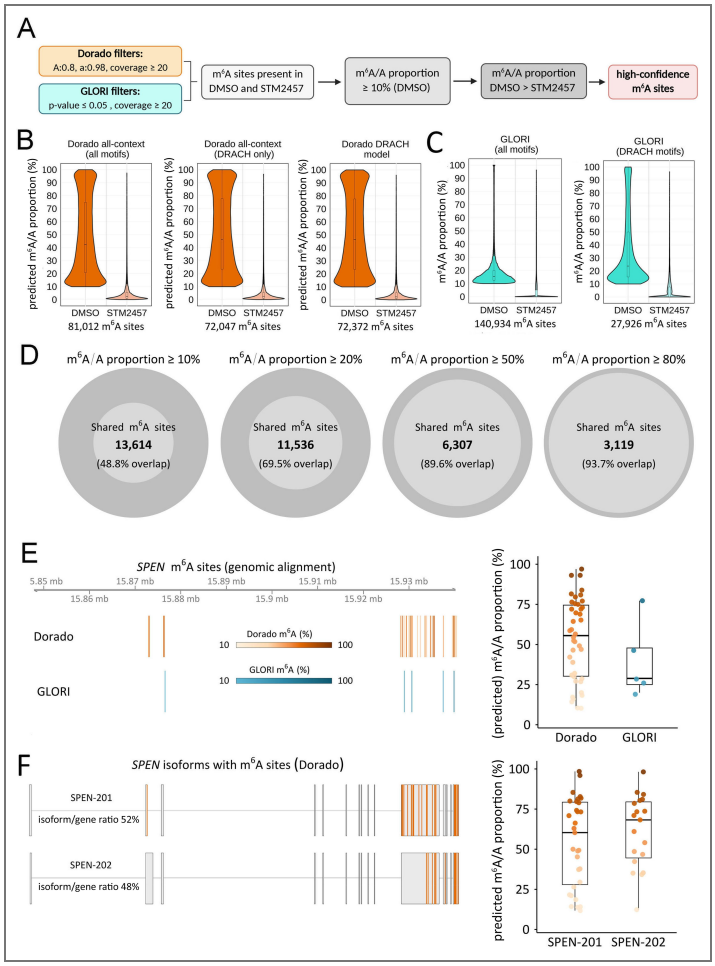


Figure 3. Comparison of m A sites detected by Dorado and GLORI.

(A) Filtering strategy for Dorado vs GLORI to obtain a high-confidence list of m^6A sites in NHDFs. (B) Predicted m^6A/A proportion distributions are plotted from genome-aligned DMSO- or STM2457-treated NHDF datasets using the indicated Dorado models and the filtering strategy shown in (A). (C) Predicted m^6A/A proportion distributions from GLORI performed on DMSO- or STM2457-treated NHDFs are plotted, using the filtering strategy shown in (A). In addition, m^6A sites were filtered for DRACH motifs (right plot). (D) Overlap of high confidence m^6A sites detected by Dorado vs GLORI. GLORI and Dorado datasets were filtered using increasing predicted m^6A/A proportion cutoffs (from left to right). GLORI sites overlapping with Dorado-identified sites are shown in light grey, whereas sites unique to GLORI are shown in darker grey. (E) m^6A sites in the *SPEN* gene detected by GLORI and Dorado. Boxplots (right) represent the predicted m^6A/A proportions per m^6A site shown on the left. (F) *SPEN* isoforms with m^6A sites, determined by Dorado. Boxplots (right) show the distribution of predicted m^6A/A proportions on the m^6A sites on SPEN-201 and SPEN-202.

context calling permits m^{6}A prediction at all adenosines, whereas DRACH-context calling restricts predictions to adenosines within DRACH motifs (DRACH, where D denotes A, G or U; R denotes A and G and H denotes A, C or U). Motif analysis of sites identified using all-context calling revealed nearly 9,000 sites outside DRACH motifs. Accordingly, when all-context calls were restricted to DRACH motifs, the resulting number of sites was similar to that obtained using DRACH-context calling (Fig. 3B). Unexpectedly, GLORI detects a substantially higher number of sites when using all motifs (140,934) compared to DRACH motifs alone (27,926) (Fig. 3C). The predominance of non-DRACH sites among the additional GLORI calls may reflect a higher contribution of false positives or background signal. Consistent with this interpretation, the overlap between Dorado and GLORI datasets remains virtually unchanged when comparing GLORI "all motif" and "DRACH motifs" (n=13,614) site sets, underscoring that the additional non-DRACH sites in GLORI (n=113,008) contributed little to the observed overlap (Fig. S21A). However, differences in detection sensitivity, sequencing depth, and other methodological characteristics of the two approaches may also contribute to the observed discrepancy and preclude attributing these additional sites solely to false-positive detection.

We obtained similar results when comparing m^{6}A sites obtained with Dorado and GLORI from HD10.6 cells. Overall, we detected a similar number of m^{6}A sites in HD10.6 cells using the all-context and DRACH-specific Dorado models (77,728 and 67,222, respectively) (Fig. S21B). By contrast, we detected 174,282 vs. 27,965 m^{6}A sites using GLORI without and with DRACH filtering, respectively (Fig. S21C).

To evaluate how predicted m^{6}A/A proportion thresholds affect the agreement between GLORI and Dorado, we compared the overlap of high-confidence GLORI-identified sites with Dorado m^{6}A (DRACH context) calls across increasing proportion cutoffs (Fig. 3D). At the lowest threshold applied (≥ 10%), 13,614 sites overlapped, representing less than half (48.8%) of GLORI sites. However, as the stringency increased, the relative overlap increased sharply: at ≥ 20%, nearly 70% of GLORI sites overlapped (11,536 sites), at ≥ 50% almost 90% overlapped (6,307 sites), and at the highest cutoff (≥ 80%), over 93% of GLORI sites (3,119 sites) were also detected by Dorado. These results demonstrate that while GLORI detects a broad set of potential m^{6}A sites, only the sites with high predicted m^{6}A/A proportions can be validated by an orthogonal approach.

GLORI relies on the high-efficiency conversion of adenosine to inosine, with overall gene A-to-I conversion rates needing to exceed 95-98% (Xu et al. 2024; Xie et al. 2025). Similarly, we observed the conversion rates for our NHDF and HD10.6 datasets to be 95% (Supplementary Table 3), but we found that small variations in gene conversion rates can strongly influence the sensitivity and specificity of GLORI, as only high-proportion m^{6}A sites overlapped with DRS-derived predictions (Fig. 3D).

DRS coupled with Dorado outperforms GLORI in isoform-specific m⁶A detection

Next, we wanted to highlight the advantage of long-read sequencing in terms of mapping reads with isoform-precision to the human transcriptome, which is not possible using GLORI due to its reliance on short-read sequencing. We first confirmed that mapping Dorado-basecalled sites to the human transcriptome results in the detection of similar numbers of m^{6}A sites in both NHDFs and HD10.6 cells (Fig. S21D,E). To assess if DRS coupled with Dorado identifies distinct m^{6}A profiles on different transcript isoforms, we analyzed the SPEN gene locus, which was previously shown to contain more than 100 m^{6}A sites (Liu et al. 2023). On a genomic level, we detected 42 m^{6}A sites using DRS, while GLORI only detected 5 m^{6}A sites (Fig. 3E). Out of the 7 total isoforms of SPEN, SPEN-201 and SPEN-202 were detectable in DRS-derived data, considering a minimal coverage of 20 (Supplementary Table 4). We found that both isoforms were expressed at equal ratios in our NHDFs. When plotting the m^{6}A sites detected by transcriptome-aligned DRS data, we found that the SPEN-201 isoform exhibited more m^{6}A sites, albeit with a lower median predicted proportion compared to the *SPEN-202* isoform (Fig. 3F). To further illustrate the ability of DRS to resolve isoform-specific m^{6}A patterns, we included three more examples in the supplementary data. The monoexonic *KRTAP1-5*, with a single detectable isoform, demonstrates that DRS identifies more

sites than GLORI, especially those with lower predicted m^{6}A/A proportions (Fig. S22A). For HSPA8, which has 5 detectable m^{6}A-containing isoforms, DRS reveals additional m^{6}A sites present in two isoforms that cannot be resolved by GLORI (Fig. S22B). Similarly, CITED2, with three detectable m^{6}A-containing isoforms, exhibits distinct isoform-specific m^{6}A modification patterns detectable only by DRS (Fig. S22C). These examples highlight that DRS is superior compared with GLORI for isoform-level m^{6}A quantification.

Dorado profiling reveals cell type differences in m A distribution and predicted proportion

To assess the reproducibility of Dorado across DRS datasets derived from different cell types, we compared the m^{6}A methylomes of NHDFs and HD10.6 cells. To ensure a fair comparison, we extracted genes that were expressed in both cell types. Next, we compared the total numbers of high confidence m^{6}A sites in NHDFs and HD10.6 cells within the genes expressed in both cell types and found that 87% of HD10.6-derived m^{6}A sites overlap with those detected in NHDFs (Fig. 4A). Site-specific predicted m^{6}A/A proportions from the shared gene set correlated strongly across cell types (Pearson r = 0.93), although we observed a global shift in the HD10.6 dataset, where predicted m^{6}A/A proportions were generally higher than in NHDFs (Fig. 4B). Plotting the m^{6}A/A proportions of the 38,217 m^{6}A sites shared between both cell types showed a strong increase in the median predicted m^{6}A/A proportions of HD10.6 cells (70.2%), compared to NHDFs (54%) (Fig. 4C).

We next aimed to conduct a comparative motif analysis of m^{6}A sites detected in NHDFs and HD10.6 cells. Sequence logos confirmed strong enrichment of the canonical DRACH consensus motif for both cell types in modification probability-filtered data, with the GGACU pentamer being the most prominent (Fig. 4D,E). Quantification of 5-mer distributions revealed highly similar motif preferences across datasets. While the global motif architecture of m^{6}A deposition is conserved between NHDFs and HD10.6 cells, we observed a trend toward higher predicted m^{6}A/A proportions in the HD10.6 dataset, particularly within the most abundant DRACH motifs (Fig. 4F,G).

Intriguingly, even after filtering, Dorado called m^{6}A at more than 100 motifs that deviate from the canonical DRACH consensus. To understand what these represent, we examined the 15 most frequently detected non-DRACH motifs in NHDFs using the DRACH model. AAAAC and AAAAA were the most abundant, followed by ACACU and GGACG. Most sites within these motifs showed a strong reduction in predicted m^{6}A/A proportions upon STM2457 treatment, indicating that they were deposited by METTL3 (Fig. S23A). Within the top 15 motifs, we found two motifs that conformed to DRAC but not DRACH. Because DRAC sites have been reported as authentic m^{6}A sites (Körtel et al. 2021; Zhou et al. 2024; Kang et al. 2025), these calls are likely biologically real (Fig. S23A). For the remaining motifs, the discrepancy that they were sensitive to STM2457 treatment but neither in a DRACH or DRAC context prompted us to ask whether these calls instead reflect systematic limitations of nanopore sequencing and basecalling. Many non-DRACH sites lie one or two nucleotides from a genuine DRACH motif. For example, m^{6}A sites within ACACU motifs are located in proximity to DRACH sites, shifted by one or two positions (Fig. S23B). Another category that contributes to the detection of non-DRACH sites are single nucleotide polymorphisms (SNPs). One example is represented by the GGACG motif, in which the final G position of the 5-mer is basecalled as A in 49% and G in 51% of all reads. This makes the sequence context a DRACH motif in half of the reads, and a non-DRACH motif in half of the other reads (Fig. S23C). Another category reflected in the non-DRACH motifs are sites that lie within homopolymer sequences. For example, several sites within AAAAA and AAAAC are reported by Dorado, but because the nanopore signal is poorly resolved within homopolymers, Dorado appears unable to localize the modified adenosine precisely, assigning it to neighboring positions rather than to the adjacent DRACH motif (Fig. S23D).

In parallel, we investigated the occurrence of non-DRACH sites using the all-context model, and as expected, Dorado predicted more than 3,000 sites. Among the top 15 motifs, GGACG, GUACU, AGACG and GGAUU were most abundant (Fig. S24A). As with the DRACH model, GGACG calls

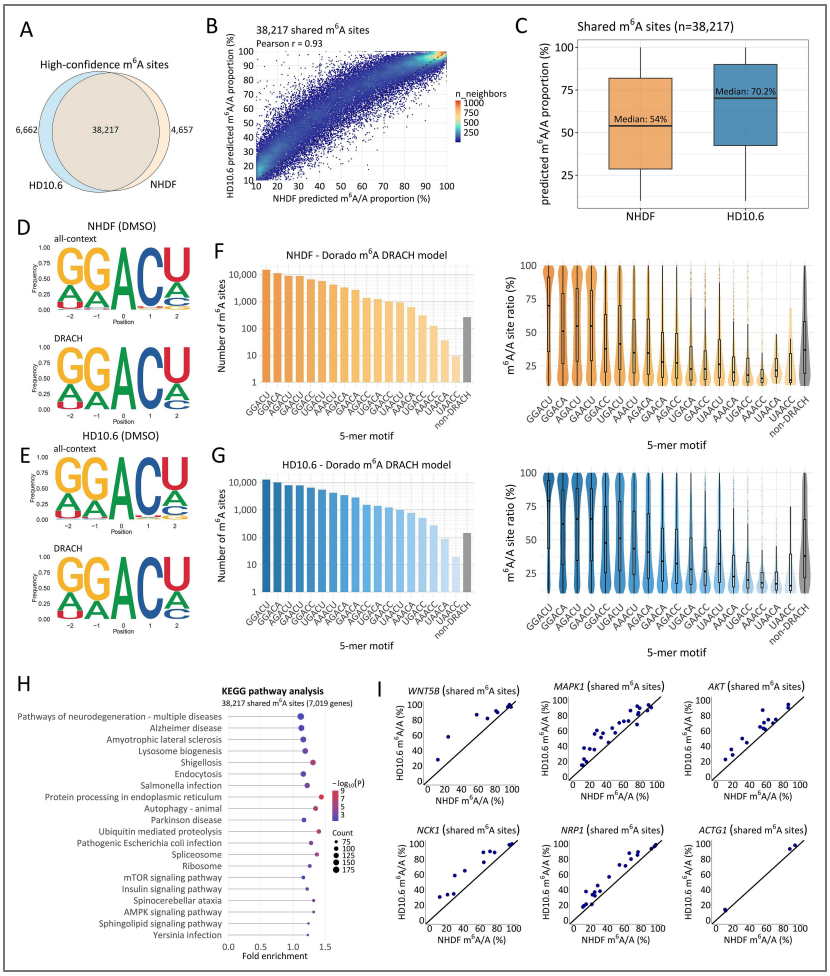


Figure 4. Comparative analysis of m⁶A modification patterns in NHDFs and HD10.6 cells using Dorado v0.9.0.

(A) Overlap of high-confidence m⁶A sites between NHDFs and HD10.6 cells. (B) Correlation of predicted m^6A/A proportions at shared m^6A sites (n=38,217). (C) Boxplots showing the predicted m^6A/A proportion distributions of the 38,217 detected in both NHDFs and HD10.6 cells. (D,E) Motif analysis showing m^6A (all-context and DRACH-context) motif logos, 5-mer distributions and predicted modification proportions in NHDFs. (F,G) Same as in (D,E) but for HD10.6 cells. (H) KEGG pathway analysis (DAVID Bioinformatics Resources) of shared m^6A-containing genes. (I) Predicted m^6A/A proportions of representative examples of individual genes from enriched pathways (H). ACTG1 serves as a housekeeping control.

derived largely from SNPs or represented genuine m^{6}A sites in a DRAC context. GGAUU, the second most abundant motif, behaved differently, as we found no evidence of miscalling or problematic sequence context. m^{6}A sites within this motif were STM2457-sensitive, suggesting authentic METTL3-dependent methylation, however, this requires further validation (Fig. S24B). Notably, the all-context calls separated into two classes: i) STM2457-sensitive sites and ii) STM2457-insensitive sites (Fig. S24A). The insensitive sites were enriched for motifs resembling the METTL16 consensus. ACAGA is a clear example, recurring in MAT2A, which is a well-established METTL16 target. Their insensitivity to METTL3 inhibition is consistent with METTL16-dependent deposition, though direct validation, for example in METTL16-depleted cells, will be required to confirm this (Fig. S24C).

Metagene analysis showed, as known previously, that m^{6}A is enriched near the stop codons and in the 3' UTRs (Fig. S25A). Several studies have examined the relationship between exon architecture and m^{6}A distribution on mRNAs (Yang et al. 2022; He et al. 2023; Luo et al. 2023), and so we also examined m^{6}A sites in monoexonic transcripts separately and found that monoexonic transcripts are more broadly m^{6}A-modified across the entire CDS (Fig. S25A). m^{6}A (both all-context and DRACH-context) distribution profiles look less typical for sites with predicted m^{6}A/A proportions lower than 10% in both NHDF and HD10.6 datasets (Fig. S25B,C). For all-context m^{6}A (Fig. S25B), the modification probability filter also proves to be deciding for the distribution profile, together with predicted m^{6}A/A proportion of ≥ 10%.

To connect shared m⁶A sites with their functional context and illustrate their distribution on individual transcripts, we conducted a KEGG pathway analysis with the genes that contain shared m^{6}A sites between NHDFs and HD10.6 cells. We used an overlap of all genes that were expressed in both cell types as a background gene set. We observed a significant enrichment in pathways related to neurodegeneration, infection, and intracellular signaling (Fig. 4H). To further illustrate these findings, we examined representative genes contained in one or more of these pathways. *WNTB5*, *MAPK1*, *AKT1*, *NCK1*, and *NRP1* all showed robust methylation in both NHDFs and HD10.6 cells, with consistently higher predicted m^{6}A/A proportions observed in the neuronal-like context. In contrast, *ACTG1*, included as a housekeeping control not represented in the enriched pathways, displayed stable and comparatively uniform methylation across both cell types (Fig. 4I).

Together, these results demonstrate that while the global architecture of m^{6}A deposition is largely conserved between primary fibroblasts and neuronal-like cells, neuronal-like cells exhibit globally higher predicted m^{6}A/A proportions, a difference that is readily detectable in individual genes. However, the biological significance of these differences remains unclear, and whether they are associated with differences in RNA abundance, stability, translation, or other cell type-specific features will require further functional investigation.

Discussion

High rates of false positives remain a major problem in epitranscriptomic mapping tools, even with the advent of modification-aware basecalling models (Cruciani et al. 2025). Comparing the outputs of orthogonal methods performed on the same input material, combined with the use of IVT RNA and/or specific methyltransferase inhibition, offers an effective mitigation strategy (Cruciani et al. 2025; Diensthuber and Novoa 2025). The results presented here establish a robust framework that minimizes false positives and enables isoform-specific and quantitative m^{6}A mapping. Through a direct comparison with GLORI, we showed that combining DRS datasets with Dorado modification-aware basecalling provides greater scalability, isoform resolution, and reproducibility across different cell types. Further, the simplicity of DRS library preparation enhances consistency, whereas GLORI is constrained by a laborious and technically challenging protocol that is extremely sensitive to adenosine conversion rates (Shen et al. 2024). Despite these limitations, concordance with Dorado increased at higher GLORI-reported m^{6}A/A proportions, with the majority of high-proportion GLORI sites also detected in our filtered DRS

datasets. By expanding our analysis across multiple human cell types, we further determine that the sites of m^6A installation are conserved for the most part but that the predicted m^6A/A proportions show cell-type-specific differences.

Crucially, the inclusion of transcriptome-wide IVT datasets and the use of a selective METTL3 inhibitor (STM2457) provide a robust approach to developing filters that dramatically decrease false-positive m^6A detection. The impact of both approaches was readily visualized by plotting modification probability distributions, enabling our determination that retaining only m^6A calls with confidence scores of 0.98 or greater retained almost all STM2457-sensitive m^6A sites while eliminating the large numbers of false-positive m^6A sites reported for the IVT datasets. This approach is supported by a recent study that highlighted that Dorado introduces large numbers of false positives outside canonical DRACH contexts (Xie et al. 2025). Our data confirm that in the unfiltered datasets, the number of false positives is higher in the all-context model compared to the DRACH model. This may reflect the degenerate nature of the canonical DRACH consensus and understudied differences in sequence preference between different cell types and organisms. Retaining only sites at which predicted m^6A/A proportions exceeded 10% further reduced false-positive rates while also focusing the results on identification of sites that are more likely to be biologically significant. The combined effect of these filters was to provide a strong agreement between the m^6A all-context and m^6A DRACH model in the detection of high-confidence m^6A sites, as observed by filtering for DRACH motifs from m^6A sites detected with the all-context model (Fig.3B).

Prior studies have emphasized the need for isoform-resolved m^6A quantification by demonstrating widespread heterogeneity in m^6A levels between genes and their isoforms (Guo et al. 2025; Xie et al. 2025). Here, we showed that aligning DRS datasets to the human transcriptome reveals differential m^6A installation patterns between isoforms, underscoring the importance of accurate, isoform-specific m^6A quantification at single-molecule resolution and highlighting the advantage of DRS against short-read sequencing approaches such as GLORI (Fig. 3E,F, Fig. S22).

Recently, the ability of Dorado and m6Anet (Hendra et al. 2022) to identify m^6A in DRS datasets from 293T and HeLa cells was systematically evaluated using IVT RNA and comparisons to published datasets obtained by GLORI and eTAM-seq (Zou et al. 2025). In this case, the authors found that Dorado (v0.7.2) outperforms m6Anet but also advocated for the importance of applying filters to reduce false-positive calls. Likewise, we found that IVT RNA served as a valuable negative control to detect false-positive calls. Furthermore, using the RNA calibration strand (RCS) as a negative control provided a convenient and simpler alternative to IVT RNA, yielding comparable probability distributions and thereby enabling a built-in control for routine experiments (Viehweger et al. 2019). While m^6A detection is generally reproducible across different Dorado modification-aware basecalling models, model-specific differences are present, i.e., model v5.0.0 produces substantially more low-confidence all-context m^6A and Ψ calls than v5.1.0 (Fig. S11C-E), highlighting that the models remain under active development and that benchmarking should be repeated when switching to a new model. Recent work further highlights that continued improvements in both nanopore sequencing chemistry and basecalling models significantly enhance RNA modification detection accuracy, emphasizing the importance of revisiting benchmarking and filtering strategies as the models evolve (Hewel et al. 2025).

Notably, Dorado v1.1.0 was included in a systematic comparison of computational approaches for the detection of six RNA modification types (Luo et al. 2026). While this study provides an important broad assessment of currently available computational methods, the workflow and filtering criteria used to derive high-confidence modification sites from Dorado predictions were not explicitly described. Our findings demonstrate that appropriate filtering and the inclusion of modification-free controls substantially reduce false-positive Dorado predictions. Thus, broader comparisons of modification-detection tools would benefit from consideration of the criteria used to define modification sites, particularly as Dorado models continue to evolve. In this respect, the framework established here complements such benchmarking efforts by providing practical guidance for obtaining high-confidence modification calls from Dorado output.

Focusing on other modifications that are detected by Dorado, we observe that stringent probability and predicted m^6A/A proportion filtering substantially reduces background predictions for inosine and Ψ, with similar filtering strategies also supported by recent studies (Verstraten et al. 2025; Ghohabi Esfahani et al. 2026). For inosine, we observed an enrichment of high-probability (> 0.99) inosine sites in genome-aligned data from STM2457-treated samples. Further studies are needed to determine whether this reflects increased intron retention or misclassification of modifications by the current inosine model. Nevertheless, very recent systematic benchmarking of Dorado modification-aware basecalling models highlights important limitations that warrant caution when interpreting these predictions (Diensthuber et al. 2026). In particular, modification-aware models were shown to exhibit cross-reactivity with chemically related modifications, including substantial cross-reactivity of the Ψ model with $m^1Ψ$ and of the m5C model with hm^5C and ac^4C. Moreover, neighboring modifications can alter the nanopore signal and contribute to false-positive predictions. Thus, even high-confidence predictions cannot necessarily be considered unambiguous evidence of the corresponding chemical modification, particularly in the absence of orthogonal validation. These findings further emphasize the importance of modification-free controls and stringent filtering, while also highlighting the need for continued improvement and benchmarking of modification-aware basecalling models.

Emerging evidence indicates that m⁶A can occur outside the DRACH consensus in vivo, and our attribution of most non-DRACH GLORI calls to false positives should not be interpreted as a claim that cellular m⁶A is restricted to this motif. Our dissection of non-DRACH DRS calls in the filtered datasets revealed that technical and biological explanations coexist. Several classes were clearly artefactual, arising from positional imprecision relative to neighboring DRACH sites, from SNPs that render the context DRACH in only a fraction of reads, or from homopolymeric stretches in which the modified adenosine cannot be localized unambiguously (Fig. S23, Fig. S24). However, we also found two abundant motifs that conformed to DRAC, but not DRACH, consistent with reports of authentic m⁶A at DRAC sites, and STM2457-sensitive GGAUU sites showed no evidence of miscalling, which reflects previous findings (Körtel et al. 2021; Zhou et al. 2024; Kang et al. 2025). Conversely, a distinct class of STM2457-insensitive sites was enriched for METTL16-like motifs, including recurrent ACAGA sites in *MAT2A* (Su et al. 2022), pointing to writers other than METTL3 rather than to error. Our results suggest that DRS is in principle well suited to detecting non-consensus m⁶A sites, as we included STM2457, which readily makes METTL3-dependent vs. -independent sites discernible. Establishing individual non-canonical sites nonetheless requires read-level inspection of the surrounding sequence, awareness of local genetic variation, and future orthogonal validation.

While DRS is the only approach that allows for robust isoform-specific m^6A mapping, comparisons with orthogonal approaches remain essential to validate findings. When comparing Dorado and GLORI using the same input RNA, we unexpectedly observed that only 50% of m^6A sites with predicted proportions of above 10% that were identified by Dorado were also present in the GLORI data, while Dorado predictions captured an increasing proportion of GLORI sites detected at high m^6A/A proportions (Fig.3D). One factor that may contribute to this discrepancy is the sub-optimal gene-wide adenosine conversion rate of approximately 95% achieved in our GLORI datasets, compared with the 98–99% reported in the original study (Shen et al. 2024). Incomplete adenosine conversion could increase background signal and potentially reduce the ability to distinguish m^6A sites with lower predicted proportions. The large number of GLORI-detected sites outside DRACH motifs may similarly reflect increased background signal (Fig. 3C), although differences in detection sensitivity, sequencing depth, and other methodological biases cannot be excluded based on the current analyses. Further methodological updates to GLORI (Sun et al. 2025) may help improve its feasibility and facilitate more robust comparisons with DRS.

Prior analyses of predicted m^6A/A proportion distributions across immortalized human cell lines, including cancer cell lines, have revealed broadly similar m^6A proportion distributions (Guo et al. 2025). By contrast, our analysis of primary fibroblasts (NHDFs) and neuronal cells (differentiated HD10.6 cells) revealed an overall increase in predicted m^6A/A proportions in neurons compared to fibroblasts, especially in genes associated with neuronal function. While

further studies are warranted to confirm and investigate the underlying mechanisms and functional consequences of these differences, this aligns with the importance of m^6A in neuronal function (Yu et al. 2021). This includes m^6A maps from synaptic and nuclear RNA isolated from the mouse dorsal hippocampus, which showed that synapse-specific RNAs had generally higher methylation levels (Sun et al. 2025). Intriguingly, neuronal regulation pathways utilize differentially spliced isoforms (Su et al. 2018), and a recent study revealed isoform-specific differences in m^6A sites and predicted m^6A/A proportions in the human brain that are linked to m^6A machinery expression and cell-type context (Gleeson et al. 2025).

In summary, using matched RNA samples from different experimental conditions and from different human cell types analyzed with both Dorado and GLORI orthogonal methodologies, we establish a robust pipeline for precise and quantitative transcriptome-wide detection of m^6A in DRS datasets. Importantly, we recommend the inclusion of a modification-free control as a best practice for accurate and reliable DRS-based RNA modification detection. Future studies should aim at extending this approach using orthogonal sequencing approaches to benchmark Dorado-based detection of Ψ, m^5C and inosine, as well as the newly released models for 2'O-methylation.

Methods

Cell culturing and STM2457 treatment

NHDFs (Lonza) were cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco), supplemented with 5% heat-inactivated FBS (HI-FBS, Gibco) and 1% penicillin-streptomycin (PS, Lonza) under standard culture conditions. The HD10.6 cell line was originally derived from fetal human dorsal root ganglia (DRG) and immortalized by overexpression of the v-myc oncogene under tetracycline regulation. Upon addition of tetracycline and neurotrophins, these cells can be differentiated into neuron-like cells that show typical neuronal morphology and nociceptive properties (Raymon et al. 1999). For our experiments, we used HD10.6 cells that have been matured for 10 days. For maintenance of undifferentiated HD10.6 cells (a gift from Anna Cliffe, University of Virginia School of Medicine), T-75 flasks were coated with 17 µg/mL fibronectin in PBS for 15 min at 37°C before adding the cell suspension. Cells were grown in Advanced DMEM/F12 (Gibco), supplemented with 1x Neurocult SM1 (STEMCELL Technologies), 2 mM L-glutamine (Gibco), 1x Primocin (Invivogen) and 10 ng/mL prostaglandin E1 (Sigma). 0.5 ng/mL bFGF (PeproTech) were added freshly to each media change. For maturation, HD10.6 cells were seeded at a density of 25.000 cells/cm² in 6-well plates which were pre-coated overnight with 50 µg/mL poly-L-ornithine hydrobromide (Sigma) in 0.5M borate buffer pH 8.5 (Boston Bioproducts), and 1 µg/mL fibronectin (overnight). After cell attachment overnight, media was changed to Neurobasal media (Gibco), supplemented with 1x Neurocult SM1, 2mM L-glutamine, 1x Primocin, 1 µg/mL doxycycline, 50 ng/mL NGF (Alomone Labs), and 25 ng/mL CNTF, GDNF and NT-3 (PeproTech). Media was changed every three days until HD10.6 cells showed complete neuronal morphology after 12 days. NHDFs and matured HD10.6 cells were treated with 30 µM STM2457 (Sigma) or DMSO for 48h, harvested in TRIzol (Invitrogen), and stored at −80°C until RNA isolation.

Cell viability testing

Cell viability was assessed using the CellTiter-Glo assay (Promega), performed according to the manufacturer's protocol. NHDF and HD10.6 cells were seeded in 96-well plates and treated with either DMSO or 30 µM STM2457 for 48h, using 10 wells per condition. Following treatment, the culture medium was aspirated and replaced with 50 µL of assay reagent. After a 2 min mixing step and a 10 min incubation at room temperature, luminescence was recorded using a PerkinElmer EnVision 2103 Multilabel Reader. Statistical significance was assessed using an unpaired, two-tailed Student's t-test.

RNA isolation

RNA was extracted by the TRIzol method according to the manufacturer's instructions. Briefly, 0.2 mL chloroform per 1 mL TRIzol were added and mixed by vortexing. After centrifugation for 30 min at 12,000xg at 4°C, the aqueous phase was transferred to isopropanol (0.5 mL per 1 mL TRIzol), mixed and incubated for 10 min at room temperature. Subsequently, the samples were centrifuged for 10 min at 12,000xg at 4°C and the RNA pellets were washed twice in 75% fresh ethanol. After air-drying for 5-10 min, RNA was dissolved in nuclease-free H_2O and stored at −80°C until further processing. One third of each sample was used for DRS and the remaining same-input material was processed for GLORI. For both DRS and GLORI, poly(A) RNA was selected using the Dynabeads™ mRNA Purification Kit (Invitrogen), according to the manufacturer's instructions.

GLORI-Sequencing

GLORI reactions were largely performed following the published protocol for RNA protection, deamination and deprotection with several minor adjustments (Shen et al. 2024). Briefly, poly(A) RNA was fragmented at 94°C for 2 min (NEB E6150S) and subjected to glyoxal protection, sodium-nitrite-mediated deamination (750 mM $NaNO_2$ at 16°C for 8h and 4°C for 6h), and deprotection in triethylammonium acetate-formamide buffer at 95°C for 10 min. Library construction was conducted using a ligation-based strategy as described by Liu and colleagues incorporating 5'adapters with 11-nt UMIs (Liu et al. 2025).

Indexed libraries were amplified with NEBNext Dual Primers Set 2 (E7780S), purified using AMPure XP beads, and validated on an Agilent TapeStation (D1000 High Sensitivity). Sequencing was performed on an Illumina NovaSeq 6000 platform (150 bp paired-end mode; > 80 million reads per sample).

For data processing, we employed the same computational pipeline described in Liu et al (Liu et al. 2025), but implemented within a Snakemake-based framework that automates the complete analysis starting from raw fastq files (Köster and Rahmann 2012). Details on software versions, workflow availability, and repository links are provided in the Data and code availability section below.

Generation of a transcriptome-wide IVT RNA library

We sequenced the IVT RNA from both NHDFs and HD10.6 cells along with the native poly(A) RNA to evaluate potential false-positive predictions by Dorado. This was achieved by converting the poly(A) RNA to cDNA using a poly(A) primer and a T7-incorporating strand switch primer and using the cDNA as a template for IVT synthesis.

Nanopore Sequencing of poly(A) RNA

250-300 ng of isolated poly(A) RNA and 25 ng RNA calibration strand (RCS) were used as input for the standard DRS SQK-RNA004 protocol. Resulting libraries were loaded onto individual PromethION RNA flow cells before sequencing on the PromethION 2 Solo instrument for 48h.

Nanopore basecalling

Basecalling was performed using Dorado versions v0.5.0, v0.6.0, v0.7.0, v0.8.0 and v0.9.0 (ONT) in super accuracy mode with modification detection enabled for m^6A (DRACH context) in v.0.5.0 and 0.6.0 (model rna004_130bps_sup@v3.0.1_m6A_DRACH@v1) and 0.8.0 and 0.9.0 (model rna004_130bps_sup@v5.1.0), m^6A (all context) in v0.7.0 (model rna004_130bps_sup@v5.0.0_m6A@v1), 0.8.0 and 0.9.0 (model rna004_130bps_sup@v5.1.0), Ψ in v0.7.0 (model rna004_130bps_sup@v5.0.0_pseU@v1), v0.8.0 and v0.9.0 (model rna004_130bps_sup@v5.1.0), and inosine and m^5C in v.0.9.0 (model rna004_130bps_sup@v5.1.0). Each dataset was processed with adapter trimming. Resulting unaligned BAM files were converted to FASTQ files containing modification tags for every read in the FASTQ read headers, using SAMtools v1.18 (Li et al. 2009) [*samtools fastq-TMM,ML*].

Genome and transcriptome level alignments

For genome-level alignment, reads were aligned against a combined human genome reference (GRCh38.p14), and the RCS gene sequence, using minimap2 v2.26 [*minimap2 -ax splice -k14 -y -uf --secondary=no*]. For transcriptome-level alignment, the reads were aligned against the human transcriptome (v47), obtained via GENCODE (Mudge et al. 2025) [*minimap2 -t 8 -ax map-ont -y -L -p 0.99*]. Using SAMtools v1.18, resulting genome- and transcriptome-aligned SAM files were converted to BAM files retaining only primary alignments [*samtools view -b -F2308 and samtools view -b -F2324, respectively*]. BAM files were sorted [*samtools sort*] and indexed [*samtools index*].

Subsampling aligned BAM files

Aligned BAM files were randomly subsampled to 200,000 reads using *reformat.sh* from BBMap v39.01 [samplereadstarget=200000]. Ten independent subsampling iterations were performed for each dataset using different random seeds (sampleseed=1-10).

Processing of RNA modification calls using Modkit

Probability distributions of the detected sites in human genome- and transcriptome-aligned reads were generated using sample-probs subcommand from Modkit v0.4.1 (Rand 2024) [sample-probs --hist --only-mapped --percentiles 0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.85,0.9 -f 0.1]. For RCS reads -f value was increased to 0.25. The resulting .tsv files were imported into RStudio (v. 2024.12.1+563) and plotted using a custom R(v. 4.4.2) script (Dorado_mod_probability_distributions.R), with the following packages: dplyr (v1.1.4), ggplot2 (v3.5.1), reshape2 (v1.4.4), patchwork (v1.3.0).

To generate extended bedMethyl files with m^6A-modified and unmodified base counts from the reads mapping against the potential modification sites, we used Modkit pileup subcommand with the thresholds [*--filter-threshold A:0.8 --filter-threshold T:0.8 --filter-threshold C:0.8 --filter-threshold G:0.8 --motif A 0 --ref $REF*] with and without modification threshold filter [e.g. *--mod-thresholds a:0.98*] to generate our "filtered" and "unfiltered" datasets, respectively. Canonical nucleotide calls (A, C, G, U) were retained with > 80% probability scores for all modification datasets. Modification thresholds were defined by our statistical framework (see below) and were set at 98% for m^6A and 99% for all the other modifications. --motif was adjusted to the base for which modification status was examined, which further restricted the analysis to only the positions that matched canonical nucleotides in the reference genome. The resulting sites were further filtered according to the read number per site (≥ 20 reads) for all the downstream analysis.

Statistical framework for modification probability cutoff determination

Modification probability distributions were analysed using a two-stage statistical framework implemented in R to identify the probability threshold above which sites are considered genuinely modified relative to an unmodified in vitro transcribed (IVT) reference. For each sample, the per-bin log2 ratio of modification probability fractions was computed as log2(sample/IVT), with a pseudocount of 1×10^{-6} added to each value prior to transformation to avoid undefined values. This ratio quantifies fold-enrichment of the sample relative to IVT at each probability bin and is insensitive to differences in absolute signal magnitude across samples. For the first stage, change-point detection was applied to the log2 ratio series using the Pruned Exact Linear Time algorithm (Killick et al. 2012) implemented in *changepoint R* package (Killick and Eckley 2014), with a manual penalty value of 1.0 and a minimum segment length of 3 bins. The first detected change-point defines the end of the baseline region (i.e. the range of probability bins over which sample and IVT distributions are statistically indistinguishable) and the start of the signal region where enrichment above baseline begins. For the second stage, the modification probability cutoff was defined as the first bin at or after the first change-point at which the log2(sample/IVT) ratio exceeded a pre-specified minimum fold-enrichment threshold. This cutoff represents the lowest modification probability at which a site shows at least X-fold more signal than would be expected

from an unmodified transcript. To characterise uncertainty in the log2 ratio estimates, bootstrap confidence intervals (95%, 2000 iterations) were computed by resampling bins with replacement and recalculating the log2 ratio distribution.

KEGG pathway enrichment analysis

KEGG pathway over-representation analysis was performed using DAVID (Huang da et al. 2009). Genes harboring shared m^{6}A sites in both NHDF and HD10.6 cells were used as input. All genes detected in both datasets were used as a custom background gene set (~13,000 genes). Pathways were considered significant at a P-value ≤ 0.05. For visualization, the top 20 KEGG pathways containing the highest number of genes were selected.

Analysis and visualization of predicted modification proportions and distributions

Histograms of predicted modification proportion distributions, as well as metaplots, were generated using RStudio and R v4.4.2 with the following libraries: data.table, patchwork, scales, tidyr, dplyr, ggplot2[4]. Access to scripts used for analysis and visualization is detailed below.

Supplementary figures

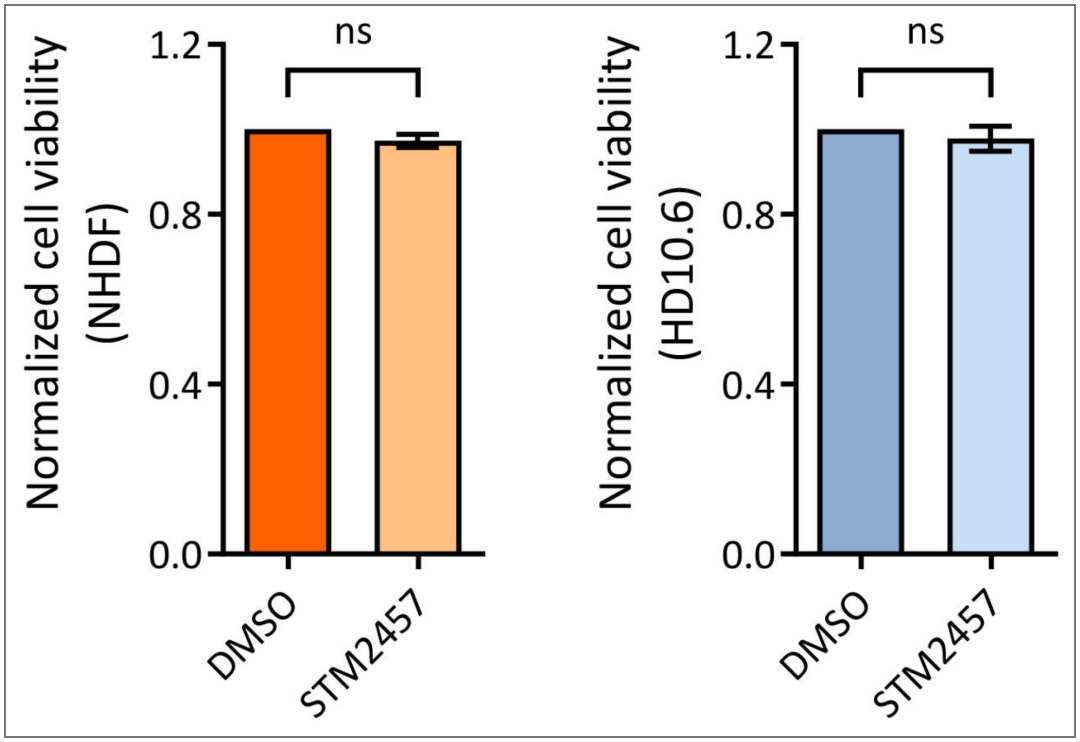


Figure S1. Cell viability assays in STM2457-treated NHDFs and HD10.6 cells. Cell viability was assessed in NHDF and HD10.6 cells after 48h of DMSO or 30 μM STM2457 treatment using CellTiter-Glo (Promega). Statistical significance was determined from 10 replicate wells per condition using an unpaired, two-tailed Student's t-test.

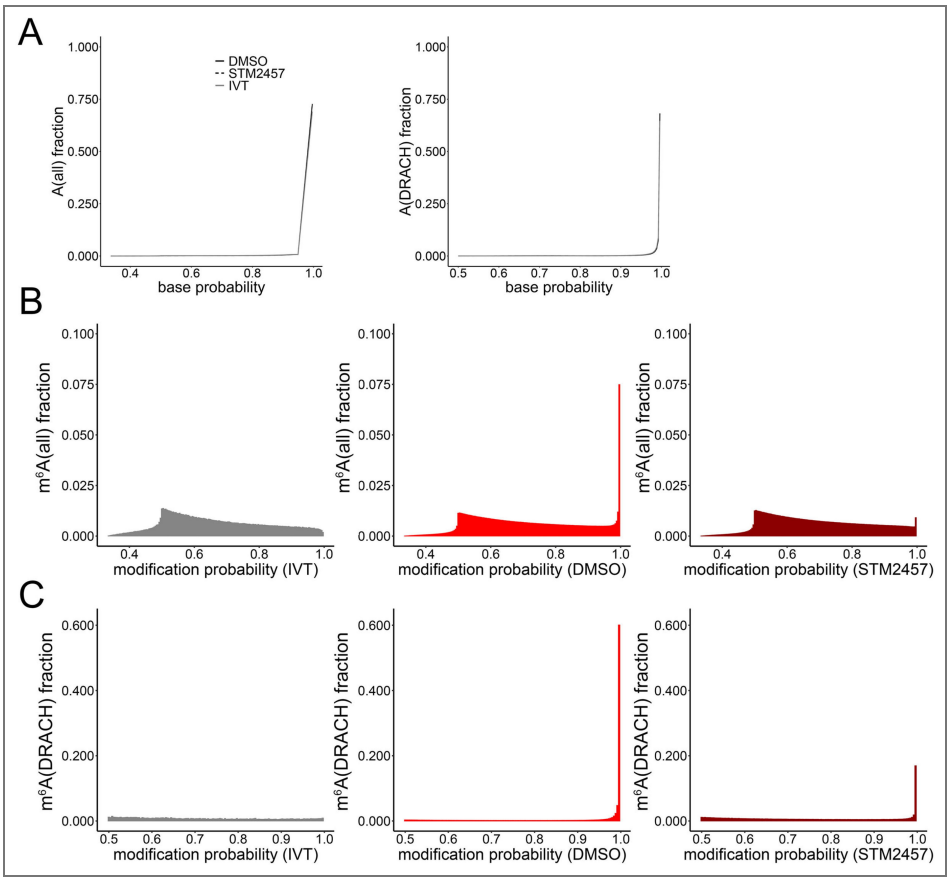


Figure S2. Evaluation of Dorado by base and modification probability distributions.

(A) Dorado v0.9.0-generated basecall accuracy probability values for A (all-context and DRACH-context basecaller models) from genome-aligned DMSO- and STM2457-treated NHDF samples against IVT poly(A) NHDF RNA. A probability score distribution is shown as a black solid line for DMSO, STM2457 – black dashed line, IVT – grey solid line. (B) Dorado v0.9.0-generated modification probability distributions, visualized as histograms, for all-context m^6A from genome-aligned DMSO- and STM2457-treated NHDFs and IVT NHDFs. IVT modification probability score distribution is shown in grey, DMSO – bright red, STM2457 – dark red. (C) Same as in (B), but for DRACH-context m^6A sites.

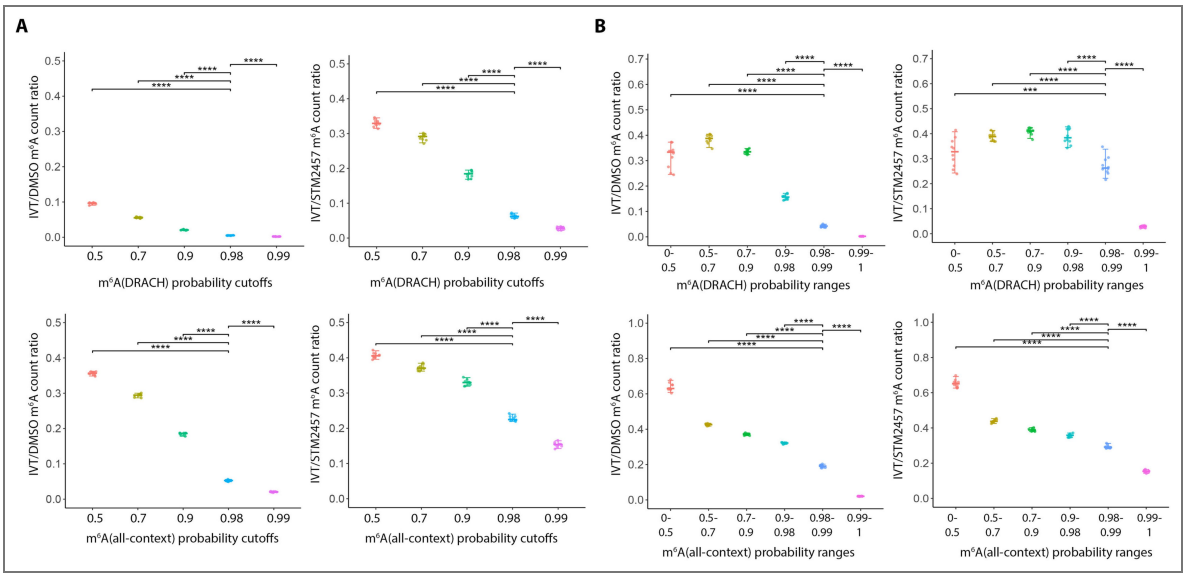


Figure S3. IVT-based estimation of false-positive m A detection across modification probability thresholds and ranges in NHDFs.

(A) False-positive m^{6}A ratios across increasing modification probability cutoffs. For each cutoff (0.5, 0.7, 0.9, 0.98, and 0.99), the number of m^{6}A calls detected in the IVT control was divided by the corresponding number detected in DMSO- or STM2457-treated NHDF samples. (B) False-positive m^{6}A ratios calculated for discrete modification probability ranges (0–0.5, 0.5–0.7, 0.7–0.9, 0.9–0.98, 0.98–0.99, and 0.99–1). In both panels, the upper row shows DRACH-context m^{6}A calls and the lower row shows all-context m^{6}A calls; left and right plots show ratios relative to DMSO- and STM2457-treated samples, respectively. Each point represents an independently generated subsample of 200,000 reads. Statistical significance was assessed using pairwise comparisons derived from linear mixed-effects models, with sample included as a random effect and Benjamini–Hochberg correction for multiple comparisons. ****P < 0.0001.

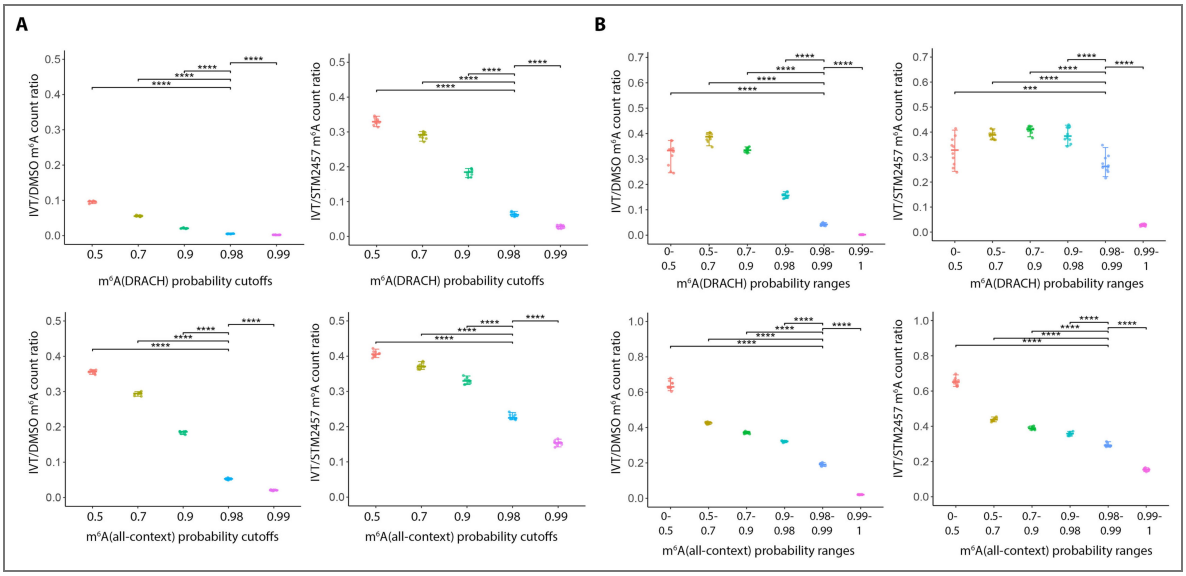


Figure S4. IVT-based estimation of false-positive m A detection across modification probability thresholds and ranges in HD10.6 cells.

(A) False-positive m^{6}A ratios across increasing modification probability cutoffs. For each cutoff (0.5, 0.7, 0.9, 0.98, and 0.99), the number of m^{6}A calls detected in the IVT control was divided by the corresponding number detected in DMSO- or STM2457-treated NHDF samples. (B) False-positive m^{6}A ratios calculated for discrete modification probability ranges (0–0.5, 0.5–0.7, 0.7–0.9, 0.9–0.98, 0.98–0.99, and 0.99–1). In both panels, the upper row shows DRACH-context m^{6}A calls and the lower row shows all-context m^{6}A calls; left and right plots show ratios relative to DMSO- and STM2457-treated samples, respectively. Each point represents an independently generated subsample of 200,000 reads. Statistical significance was assessed using pairwise comparisons derived from linear mixed-effects models, with sample included as a random effect and Benjamini–Hochberg correction for multiple comparisons. ****P < 0.0001.

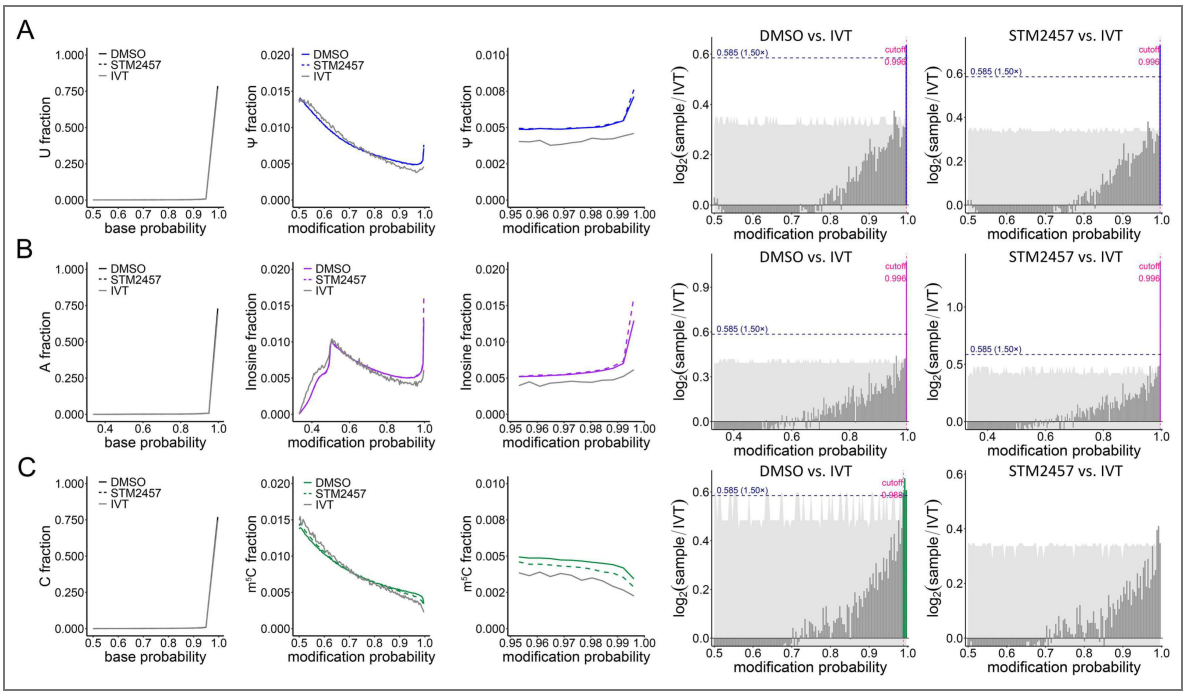


Figure S5. Modification probability cutoffs are necessary to identify Ψ and inosine sites more accurately.

(A) (left) Dorado v0.9.0-generated basecall accuracy probability values for U (left) and modification probability distributions for Ψ (right) from genome-aligned DMSO- and STM2457-treated NHDF samples are plotted against IVT poly(A) RNA. Ψ probabilities are visualized in blue solid and dashed lines for DMSO and STM2457, respectively and in grey solid line for IVT. (right) Bar plots showing the per-bin log2(sample/IVT) ratio of modification probability fractions for the DMSO- and STM2457-treated NHDFs conditions relative to the unmodified IVT reference. The horizontal navy dashed line marks the fold-enrichment threshold used to call enrichment. Bars are coloured by significance relative to the fold-enrichment threshold (blue, log2 ≥ 1.5-fold; dark grey, not significant). The light grey ribbon indicates the 95% bootstrap confidence interval around the log2 ratio estimate at each bin. The vertical pink dashed line indicates the derived modification probability cutoff above which sites are considered genuinely modified. (B) Same as in (A) but showing A base accuracy probabilities and inosine sites. Inosine probabilities are visualized in purple solid and dashed lines for DMSO and STM2457, respectively and in grey solid line for IVT. (C) Same as in (A) but showing C base accuracy probabilities and m^5C sites. m^5C probabilities are visualized in green solid and dashed lines for DMSO and STM2457, respectively and in grey solid line for IVT.

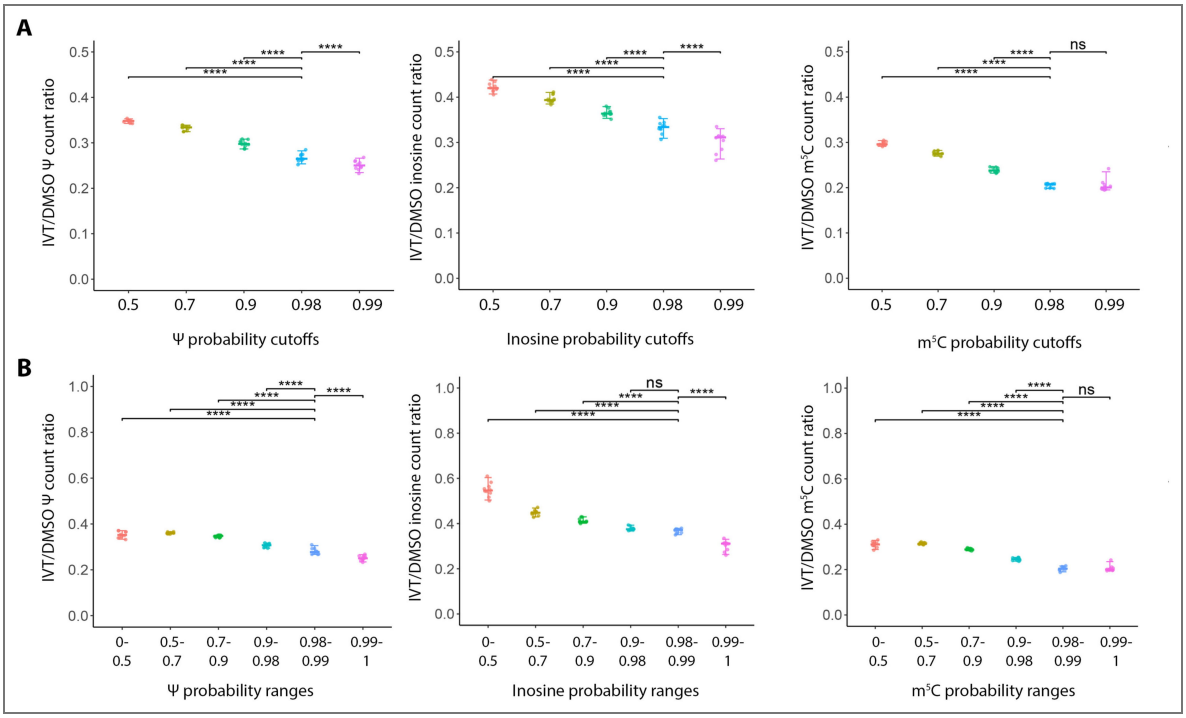


Figure S6. IVT-based estimation of false-positive Ψ, inosine, and m C detection across modification probability thresholds and ranges in NHDFs.

(A) False-positive modification ratios across increasing modification probability cutoffs (0.5, 0.7, 0.9, 0.98, and 0.99) for Ψ, inosine, and m^5C. For each cutoff, the number of modification calls detected in the IVT control was divided by the corresponding number detected in DMSO-treated NHDF samples. (B) False-positive modification ratios calculated for discrete modification probability ranges (0–0.5, 0.5–0.7, 0.7–0.9, 0.9–0.98, 0.98–0.99, and 0.99–1). From left to right, plots show results for Ψ, inosine, and m^5C. Each point represents an independently generated subsample of 200,000 reads. Statistical significance was assessed using pairwise comparisons derived from linear mixed-effects models, with sample included as a random effect and Benjamini–Hochberg correction for multiple comparisons. ****P < 0.0001; ns, not significant.

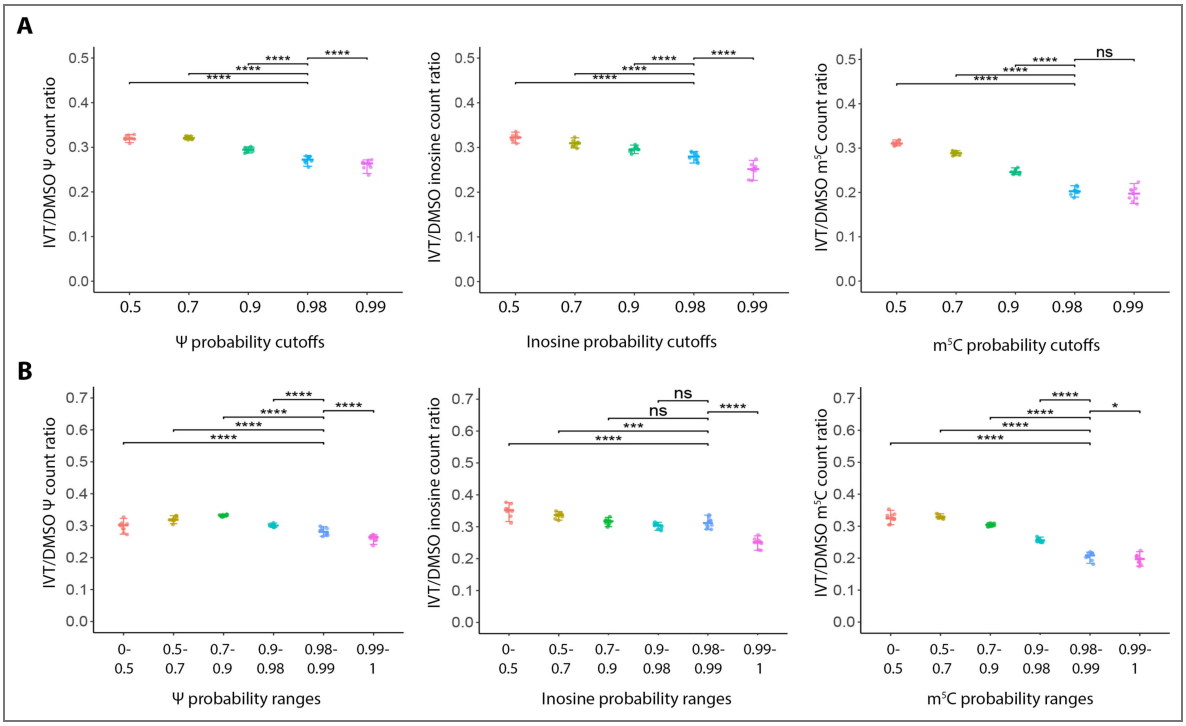


Figure S7. IVT-based estimation of false-positive Ψ, inosine, and m C detection across modification probability thresholds and ranges in HD10.6 cells.

(A) False-positive modification ratios across increasing modification probability cutoffs (0.5, 0.7, 0.9, 0.98, and 0.99) for Ψ, inosine, and m^5C. For each cutoff, the number of modification calls detected in the IVT control was divided by the corresponding number detected in DMSO-treated HD10.6 samples. (B) False-positive modification ratios calculated for discrete modification probability ranges (0–0.5, 0.5–0.7, 0.7–0.9, 0.9–0.98, 0.98–0.99, and 0.99–1). From left to right, plots show results for Ψ, inosine, and m^5C. Each point represents an independently generated subsample of 200,000 reads. Statistical significance was assessed using pairwise comparisons derived from linear mixed-effects models, with sample included as a random effect and Benjamini–Hochberg correction for multiple comparisons. ****P < 0.0001; ns, not significant.

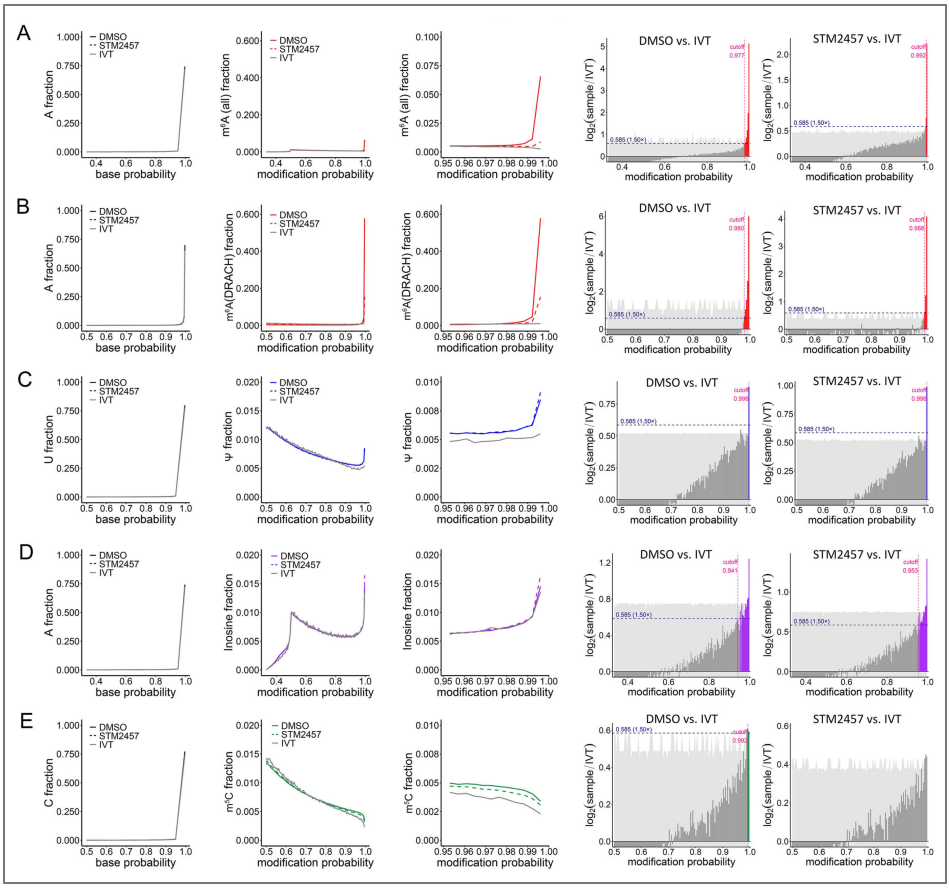


Figure S8. Modification probability cutoffs are necessary to identify modification sites accurately, shown for transcriptome-aligned data.

Dorado v0.9.0-generated basecall accuracy probability values for A, U and C (left, A-E) and modification probability distributions for m^6A (all-context), m^6A (DRACH-context), Ψ, inosine and m^5C (middle, A-E) in transcriptome-aligned DMSO- and STM2457-treated NHDFs, plotted against the values from IVT NHDF poly(A) RNA. (right) Bar plots showing the per-bin log2(sample/IVT) ratio of modification probability fractions for the DMSO- and STM2457-treated NHDFs conditions relative to the unmodified IVT reference. The horizontal navy dashed line marks the fold-enrichment threshold used to call enrichment. Bars are coloured by significance relative to the fold-enrichment threshold (blue, log2 ≥ 1.5 fold; dark grey, not significant). The light grey ribbon indicates the 95% bootstrap confidence interval around the log2 ratio estimate at each bin. The vertical pink dashed line indicates the derived modification probability cutoff above which sites are considered genuinely modified.

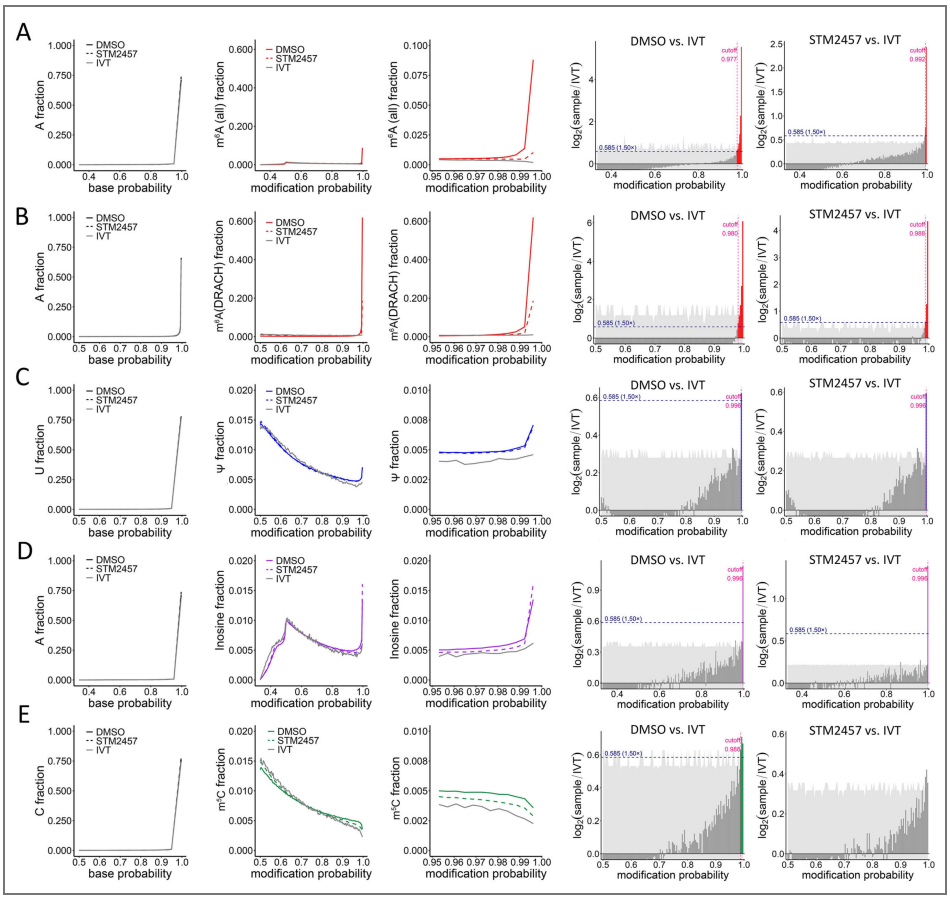


Figure S9. Modification probability distribution profiles in genome-aligned DMSO-and STM2457-treated HD10.6 cells.

Dorado v0.9.0-generated basecall accuracy probability values for A, U and C (left, A-E) and modification probability distributions for m^6A (all-context), m^6A (DRACH-context), Ψ, inosine and m^5C (middle, A-E) in human genome-aligned DMSO- and STM2457-treated HD10.6 cells, plotted against the values from IVT HD10.6 poly(A) RNA. (right) Bar plots showing the per-bin log2(sample/IVT) ratio of modification probability fractions for the DMSO- and STM2457-treated HD10.6 conditions relative to the unmodified IVT reference. The horizontal navy dashed line marks the fold-enrichment threshold used to call enrichment. Bars are coloured by significance relative to the fold-enrichment threshold (blue, log2 ≥ 1.5-fold; dark grey, not significant). The light grey ribbon indicates the 95% bootstrap confidence interval around the log2 ratio estimate at each bin. The vertical pink dashed line indicates the derived modification probability cutoff above which sites are considered genuinely modified.

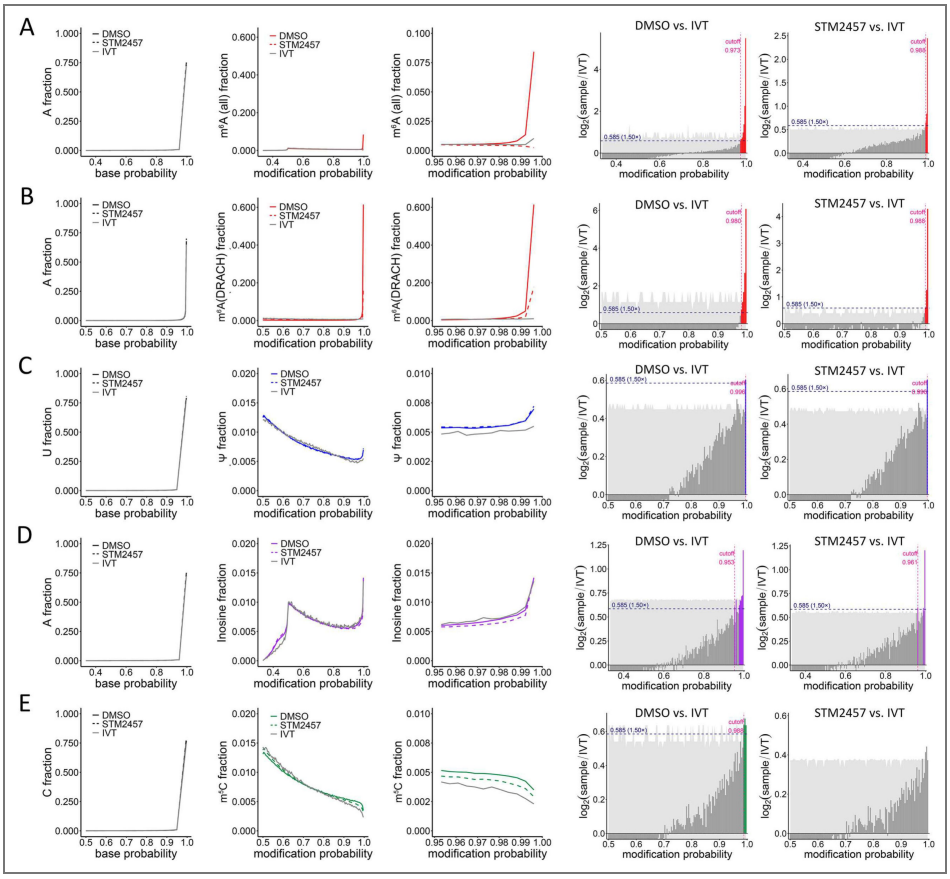


Figure S10. Modification probability distribution profiles in transcriptome-aligned DMSO-and STM2457-treated HD10.6 cells.

Dorado v0.9.0-generated basecall accuracy probability values for A, U and C (left, A-E) and modification probability distributions for m^6A (all-context), m^6A (DRACH-context), Ψ, inosine and m^5C (middle, A-E) in human transcriptome-aligned DMSO-and STM2457-treated HD10.6 cells, plotted against the values from IVT HD10.6 poly(A) RNA. (right) Bar plots showing the per-bin log2(sample/IVT) ratio of modification probability fractions for the DMSO- and STM2457-treated HD10.6 conditions relative to the unmodified IVT reference. The horizontal navy dashed line marks the fold-enrichment threshold used to call enrichment. Bars are coloured by significance relative to the fold-enrichment threshold (blue, log2 ≥ 1.5-fold; dark grey, not significant). The light grey ribbon indicates the 95% bootstrap confidence interval around the log2 ratio estimate at each bin. The vertical pink dashed line indicates the derived modification probability cutoff above which sites are considered genuinely modified.

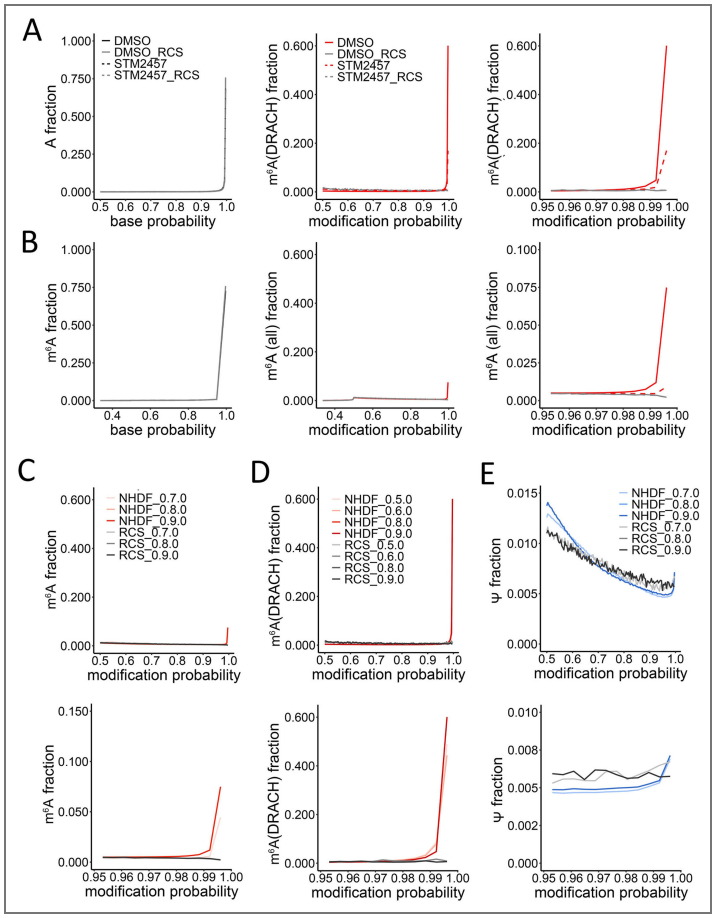


Figure S11. Modification site probabilities detected with different versions and models of Dorado and compared to RCS.

(A,B) DRACH-context and all-context m^{6}A modification probability distributions from genome-aligned DMSO (solid red line)- and STM2457 (dashed red line)-treated NHDFs are plotted against RCS negative control (grey). (C) Modification probability distributions for all-context m^{6}A (Dorado v0.7.0 (model v5.0.0), v0.8.0 (model v5.1.0) and v0.9.0 (model v5.1.0)), the bottom panel showing the distributions in 0.95-1.00 modification range. (D) Same as in (C), but for DRACH-context m^{6}A (Dorado v0.5.0 (model v3.0.1), v0.6.0 (model v3.0.1), v0.8.0 (model v5.1.0) and v0.9.0 (model v5.1.0)). (E) Same as in (C), but for Ψ (Dorado v0.7.0 (model v5.0.0), v0.8.0 (model v5.1.0), v0.9.0 (model v5.1.0)). All the plots in C-E are generated from genome-aligned DMSO-treated NHDF datasets and are plotted against RCS negative controls.

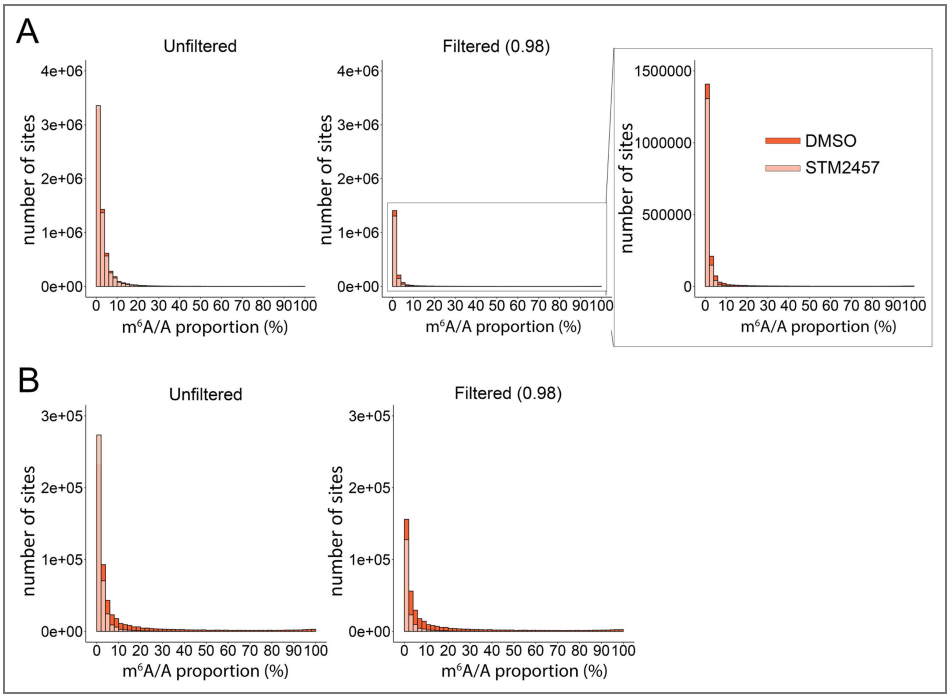


Figure S12. Evaluation of Dorado by the distributions of predicted m A/A proportions.

(A) Distributions of predicted m^6A/A proportions for all-context m^6A from DMSO-(in red) or STM2457-treated (in pink) NHDF datasets (genome-aligned and filtered for ≥ 20 reads) are plotted, comparing 0.98 modification probability-unfiltered (left) and filtered (right) sites. (B) Same as in (A), but for DRACH-context m^6A sites.

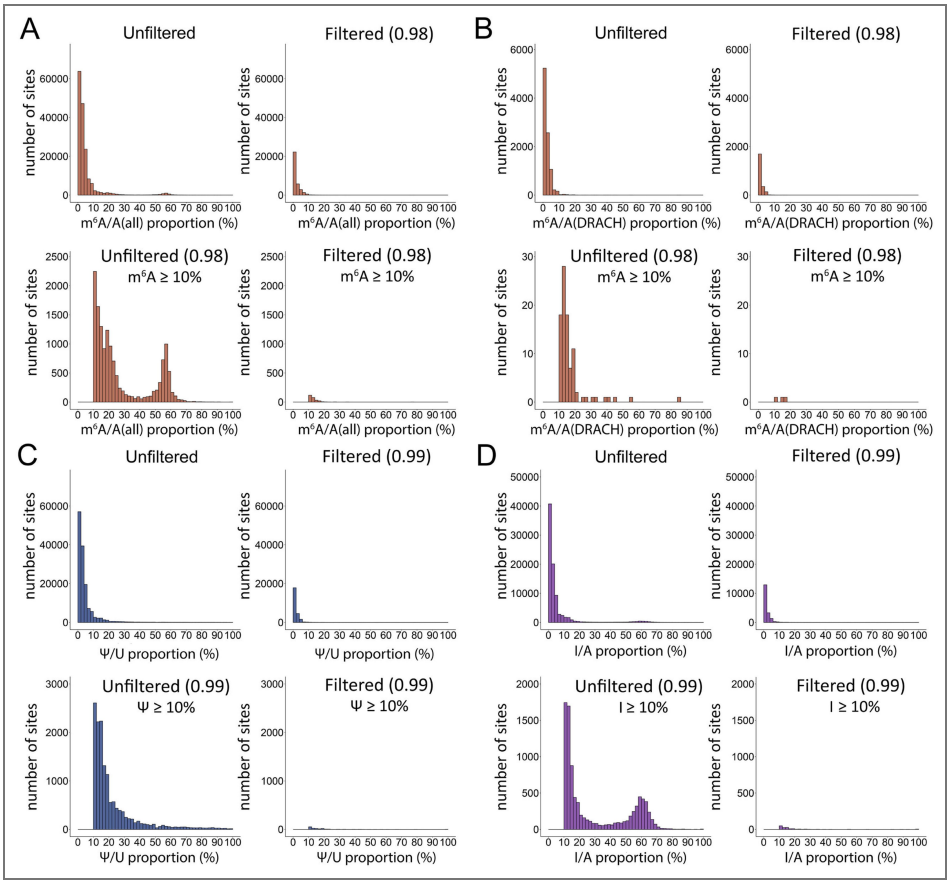


Figure S13. Distributions of predicted m A/A, Ψ/U and inosine/A proportions in IVT datasets.

(A) All-context predicted m^6A/A proportion distributions (all proportions (top) and ≥ 10% proportions (bottom)) in (genome-aligned) IVT RNA from DMSO-treated NHDFs, comparing modification probability-unfiltered and filtered data. The bottom panel shows distributions only for ≥ 10% proportion sites. (B) As in (A), but for DRACH-context m^6A. (C) As in (A), but for Ψ. (D) As in (A), but for inosine.

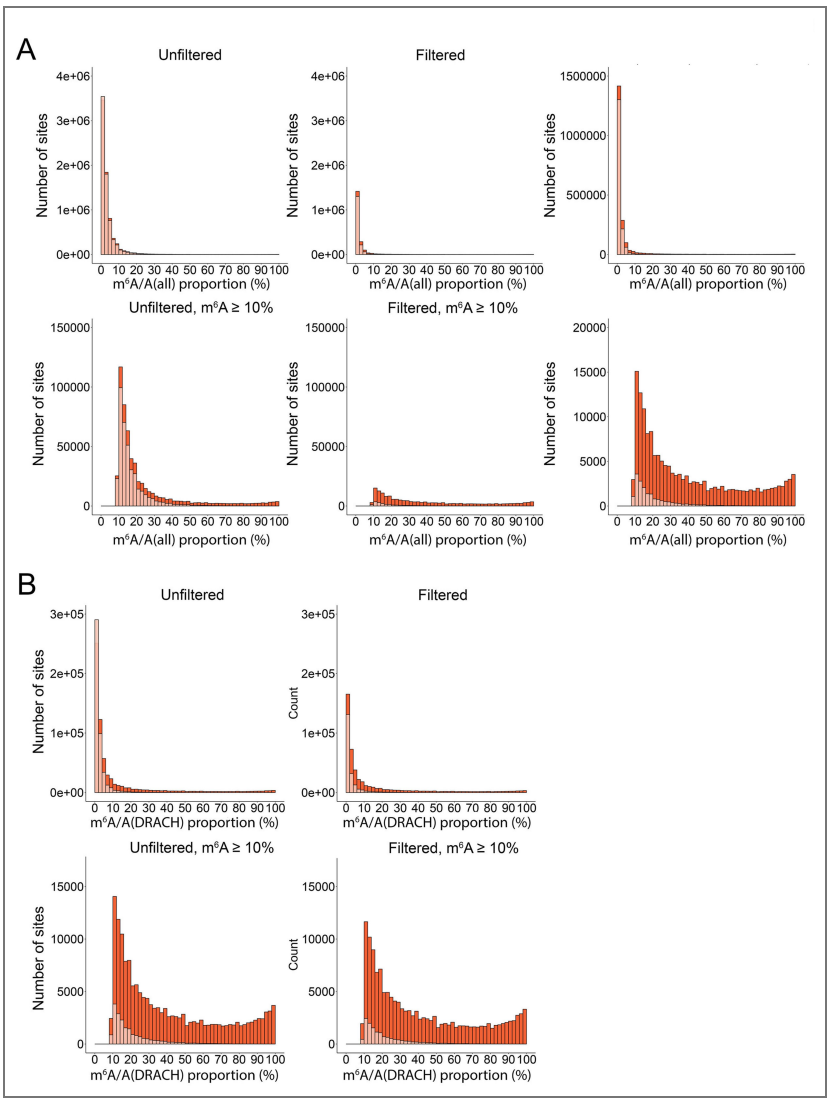


Figure S14. Evaluation of Dorado by predicted modification proportion distributions in transcriptome-aligned NHDFs.

(A) Distributions of predicted m^{6}A/A proportions for all-context m^{6}A from DMSO-(in red) or STM2457-treated (in pink) NHDF datasets (transcriptome-aligned and filtered for ≥ 20 reads) are plotted, assessing the 0.98 modification probability cutoff. The bottom panel shows distributions only for ≥ 10%-proportion sites. (B) Same as in (A), but for DRACH-context m^{6}A sites.

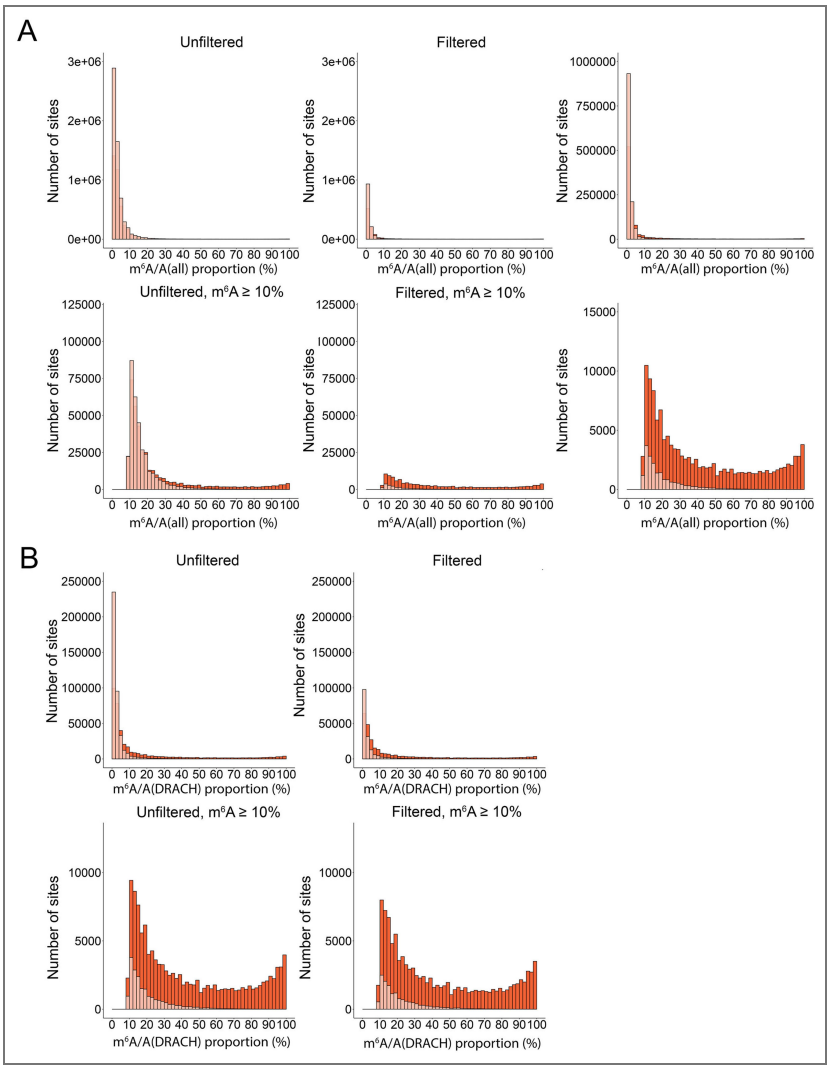


Figure S15. Evaluation of Dorado by distributions of predicted m A/A proportions in transcriptome-aligned HD10.6 cells.

(A) Distributions of predicted m^{6}A/A proportions for all-context m^{6}A from DMSO-(in red) or STM2457-treated (in pink) HD10.6 datasets (transcriptome-aligned and filtered for ≥ 20 reads) are plotted, assessing the 0.98 modification probability cutoff. The bottom panel shows distributions only for ≥ 10%-proportion sites. (B) Same as in (A), but for DRACH-context m^{6}A sites.

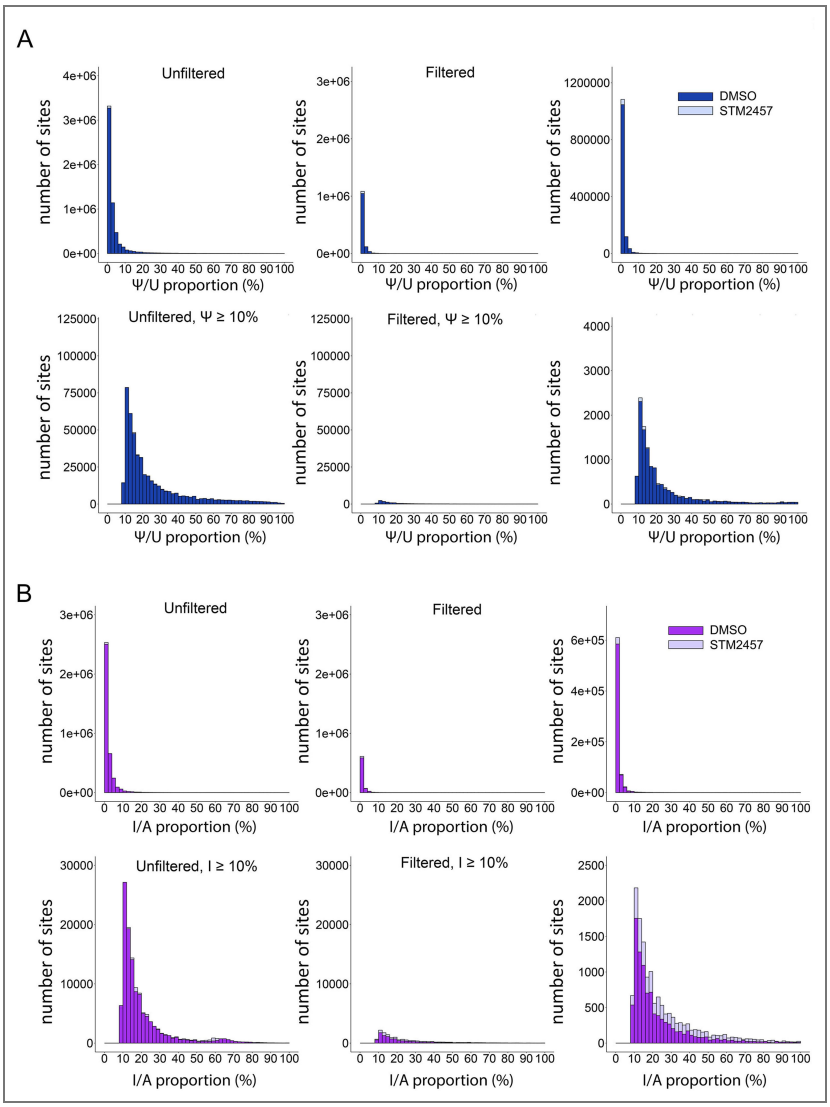


Figure S16. Distributions of predicted Ψ/U and inosine/A proportions in NHDFs.

(A) Distributions of all (left) and ≥ 10% (right) predicted Ψ/U proportions in the 0.99 modification probability-unfiltered and filtered data from genome-aligned DMSO-(darker color) and STM2457-treated (lighter color) NHDFs (filtered for ≥ 20 reads). (B) Same as in (A), but for inosine.

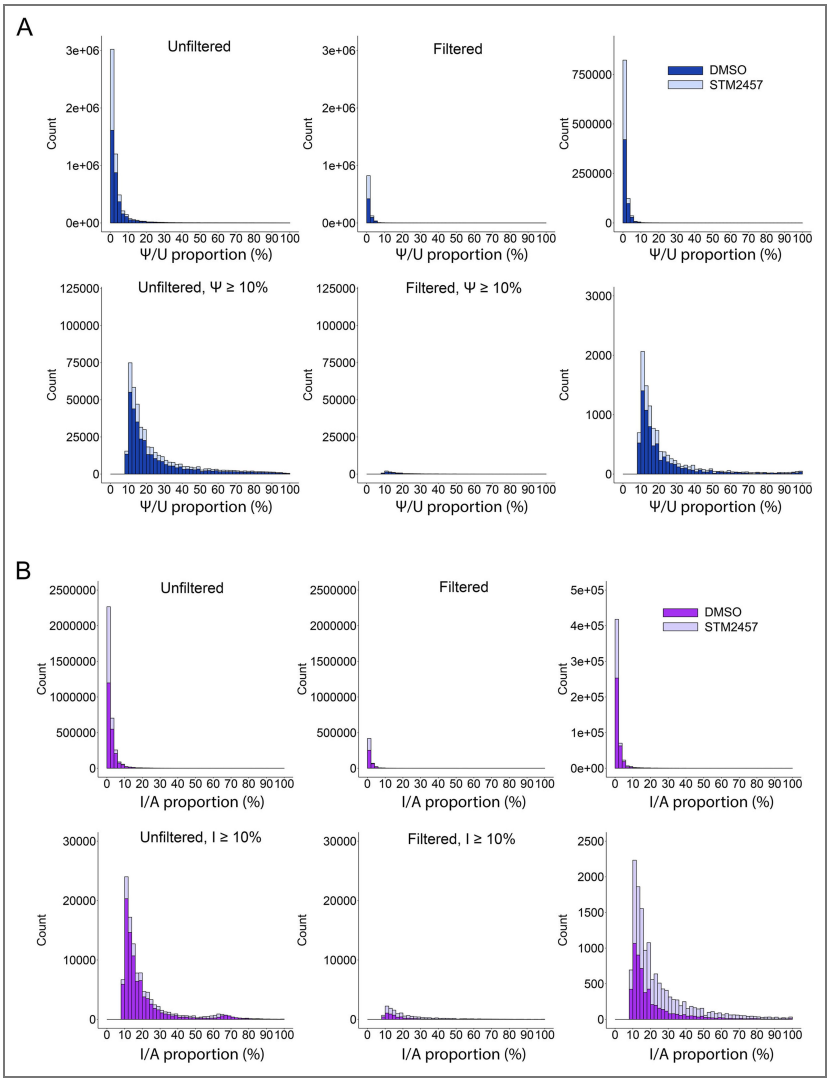


Figure S17. Distributions of predicted Ψ/U and inosine/A proportions in HD10.6 cells.

(A) Distributions of all and ≥ 10% predicted Ψ/U proportions in the 0.99 modification probability-unfiltered and filtered data from genome-aligned DMSO-(darker color) and STM2457-treated (lighter color) HD10.6 cells (filtered for ≥ 20 reads). (B) Same as in (A), but for inosine.

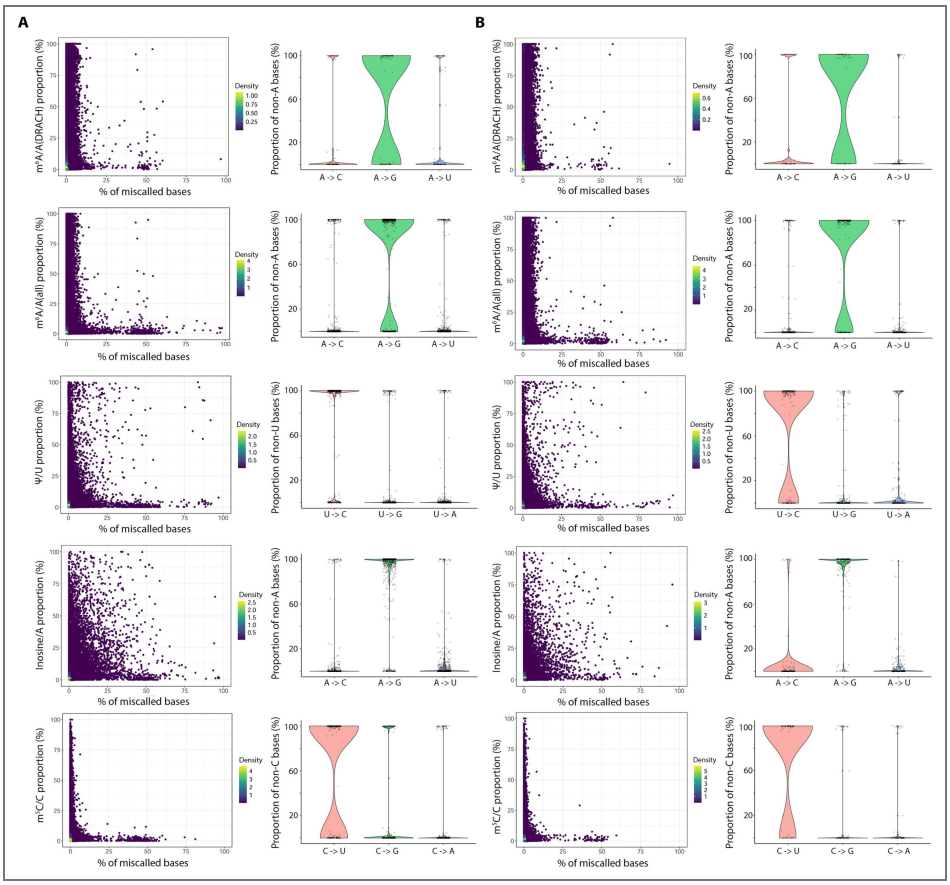


Figure S18. Relationship between predicted RNA modification proportions and nucleotide basecalling errors in NHDF and HD10.6 datasets.

(A) NHDF and (B) HD10.6 datasets showing, from top to bottom, DRACH-context m^6A, all-context m^6A, Ψ, inosine, and m^5C sites. Prior to analysis, sites were filtered using modification probability thresholds of ≥0.98 for m^6A and ≥0.99 for Ψ, inosine, and m^5C, together with a minimum valid-read coverage of 20. For each modification type, scatterplots show the predicted modification proportion generated by Dorado and the percentage of overlapping reads basecalled as a nucleotide other than the corresponding reference nucleotide (A for m^6A and inosine, T for Ψ, and C for m^5C). Each point represents an individual modification site, with point color indicating density. Adjacent violin plots characterize the identity of alternative basecalls at sites for which ≥25% of all overlapping reads were basecalled as a non-reference nucleotide. For each site, the proportion of mis-basecalled reads assigned to each of the three alternative nucleotides was calculated, with individual points representing individual modification sites. Violin distributions therefore show the relative contribution of each alternative nucleotide among mis-basecalled reads at sites exhibiting high levels of basecalling mismatch.

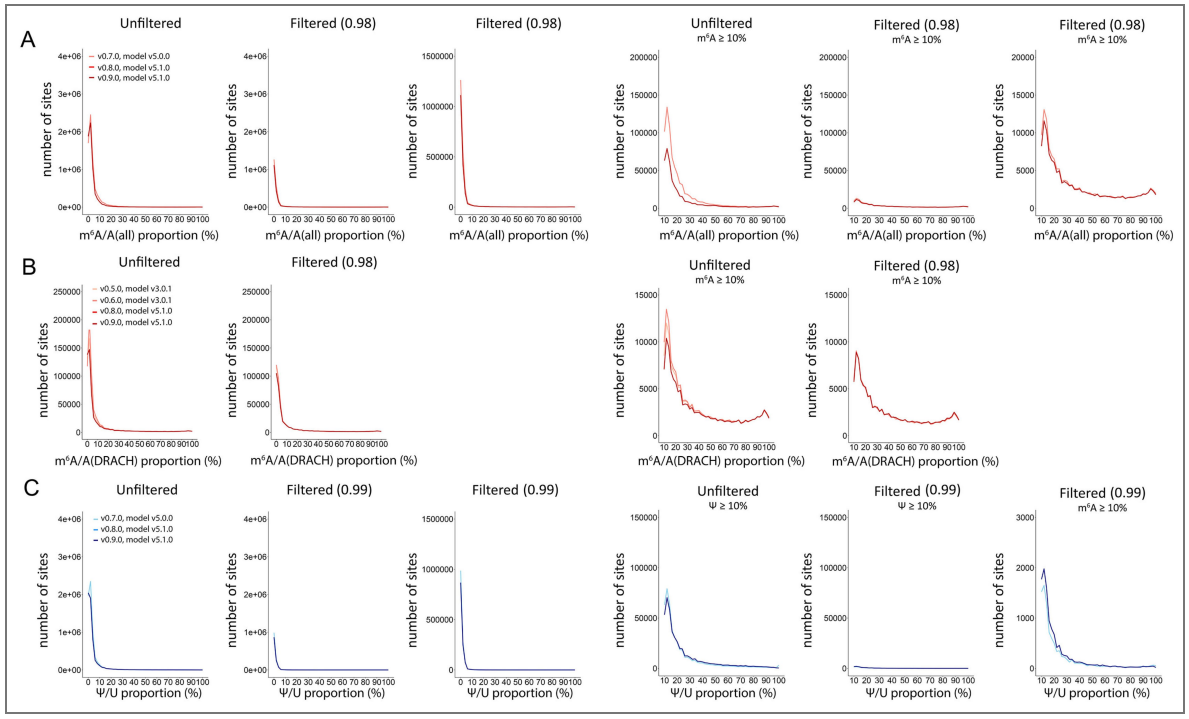


Figure S19. Predicted m A/A and Ψ/U proportion distributions in different versions of Dorado.

Distributions of predicted modification proportions (all – left, ≥ 10% – right) in a line graph format for (A) all-context m^6A detected in genome-aligned DMSO-treated NHDF dataset, with Dorado versions 0.7.0 (model v5.0.0), 0.8.0 (model v5.1.0) and 0.9.0 (model v5.1.0) colored in gradients of red, darkness increasing with the later versions. (B) DRACH-context m^6A detected with Dorado versions 0.5.0 (model v3.0.1), 0.6.0 (model v3.0.1), 0.8.0 (model v5.1.0) and 0.9.0 (model v5.1.0), colored in gradients of red, darkness increasing with the later versions. (C) Ψ detected with Dorado versions 0.7.0 (model v5.0.0), 0.8.0 (model v5.1.0) and 0.9.0 (model v5.1.0), colored in gradients of blue, darkness increasing with the later versions. All the datasets were filtered for ≥ 20 reads.

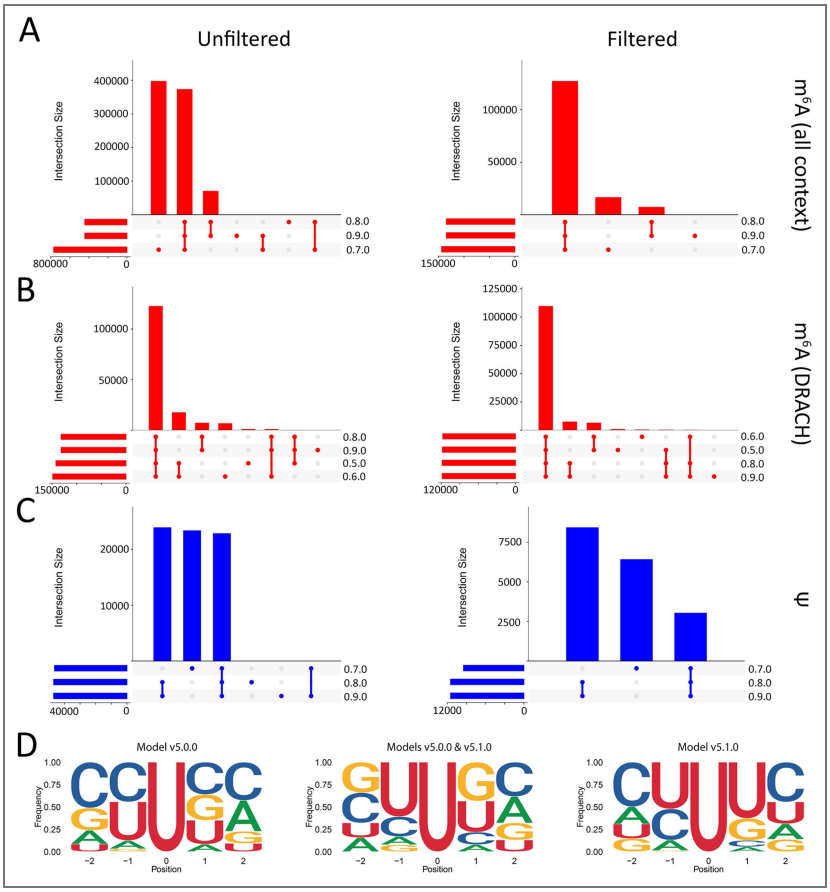


Figure S20. Comparison of modification sites detected with different versions of Dorado.

Upset plots showing unique and shared detected modifications (filtered for ≥ 20 reads, ≥ 10% predicted modification proportion) in modification probability-unfiltered and filtered genome-aligned DMSO-treated NHDF data. The versions are ordered on the plot by the number of detected sites, in ascending order (on the left). Results are shown for the following models and versions: (A) all-context m^6A detected with v.0.7.0 (model v5.0.0), v0.8.0 (model v5.1.0) and v0.9.0 (model v5.1.0), (B) DRACH-context m^6A detected with v.0.5.0 (model v3.0.1), v0.6.0 (model v3.0.1), v0.8.0 (model v5.1.0) and v0.9.0 (model v5.1.0), (C) Ψ detected with v.0.7.0 (model v5.0.0), v0.8.0 (model v5.1.0) and v0.9.0 (model v5.1.0). (D) Nucleotide frequencies in the 5-base motif of Ψ sites uniquely for model v5.0.0 (basecalled with Dorado v0.7.0), common for models v5.0.0 (basecalled with v0.7.0) and v5.1.0 (basecalled with Dorado v0.8.0 and v0.9.0), and uniquely for model v5.1.0-detected (basecalled with v0.8.0 and v0.9.0) Ψ sites. Position 0 marks the modification site.

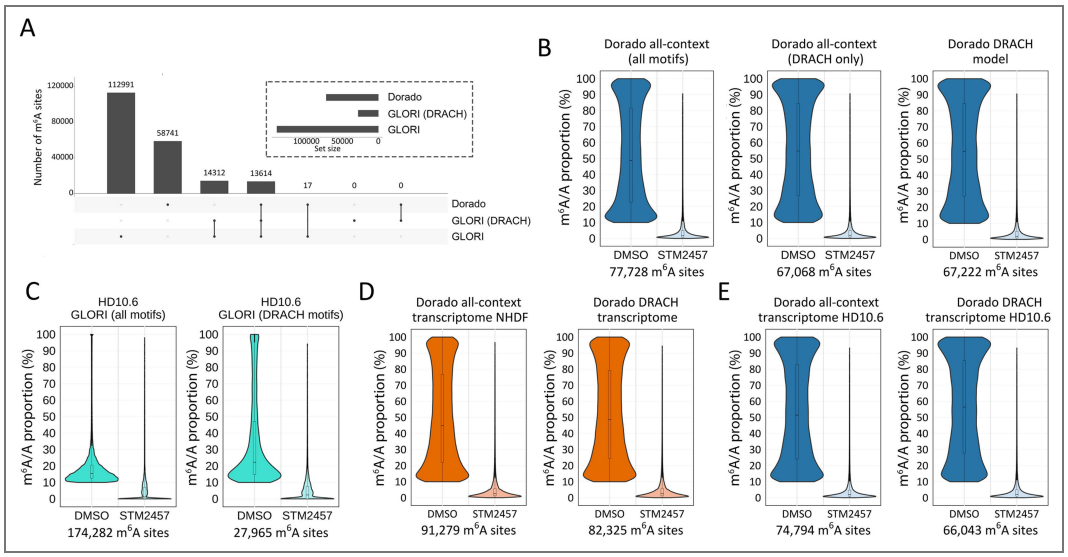


Figure S21. Comparison of m A sites detected by Dorado and GLORI-Sequencing in NHDFs and HD10.6s.

(A) Upset plot showing the number of m^6A sites detected by GLORI and Dorado in NHDFs and their overlaps. Dorado sites represent high-confidence m^6A sites obtained after filtering DRACH-model detected data. Similarly, only high confidence m^6A sites detected by GLORI are shown but were additionally filtered for DRACH sites. (B) Distributions of predicted m^6A/A proportions are plotted from genome-aligned DMSO- or STM2457-treated HD10.6 datasets using the indicated Dorado models. (C) Distributions of predicted m^6A/A proportions from GLORI performed on DMSO- or STM2457-treated NHDFs are plotted. In addition, m^6A sites were filtered for DRACH motifs (right plot). (D) Distributions of predicted m^6A/A proportions are plotted from transcriptome-aligned DMSO- or STM2457-treated NHDF datasets using the indicated Dorado models. (E) Distributions of predicted m^6A/A proportions are plotted from transcriptome-aligned DMSO- or STM2457-treated HD10.6 datasets using the indicated Dorado models.

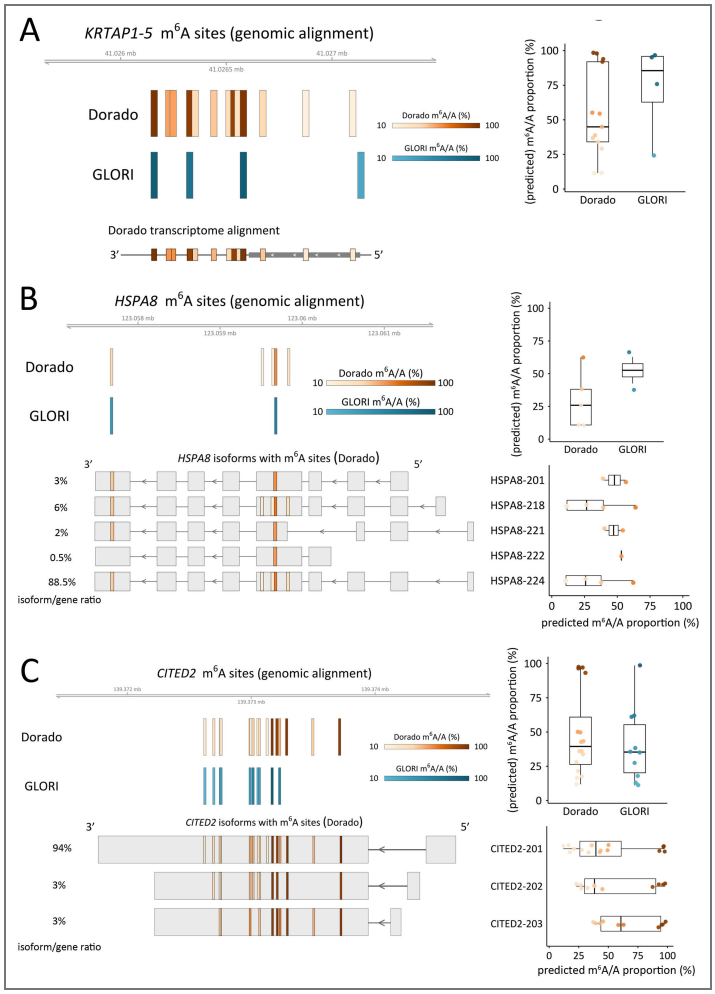


Figure S22. Isoform-level m A prediction using the Dorado DRACH model in NHDF data.

(A) m^{6}A sites in the *KRTAP1-5* gene detected by GLORI and Dorado. Boxplots (right) represent the predicted m^{6}A/A proportions per m^{6}A site shown on the left. (B) m^{6}A sites in the *HSPA8* gene detected by GLORI and Dorado. Boxplots (right) represent the predicted m^{6}A/A proportions per m^{6}A site shown on the left. Below, *HSPA8* isoforms with m^{6}A sites, determined by Dorado. Boxplots (right) show the distribution of predicted m^{6}A/A proportions on the m^{6}A sites on *HSPA8-201* and *HSPA8-202*. (C) Same as in B, but for *CITED2*.

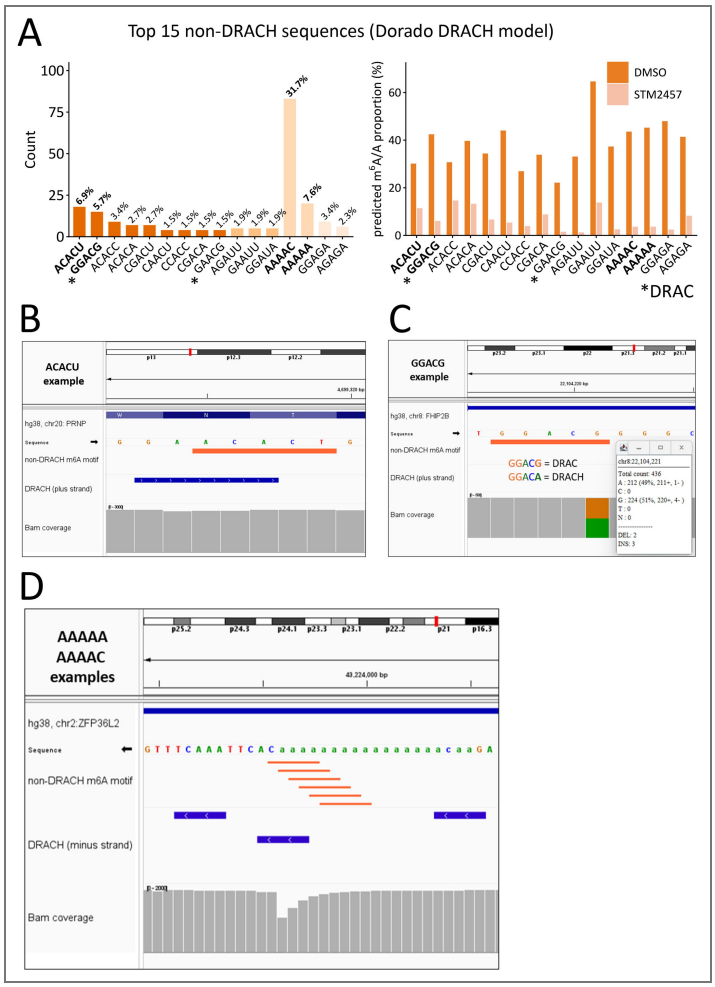


Figure S23. Non-DRACH m A calls predicted by the Dorado DRACH model in NHDF data.

(A) Top 15 most frequently detected non-DRACH motifs, ranked by occurrence (left); bars are colored by the central dinucleotide, and percentages indicate the proportion of all non-DRACH calls. Bold labels mark the 4 most abundant motifs discussed in the text. Mean predicted m A/A proportions for each motif in DMSO- and STM2457-treated cells (right). Motifs conforming to DRAC but not DRACH are indicated with asterisks. (B) Representative genome browser view of an ACACU call located one to two nucleotides from a canonical DRACH motif, with the called and adjacent positions marked. (C) Read-level basecalls at a GGACG site, where the terminal position is called as G in 51% and A in 49% of reads, rendering the context non-DRACH and DRACH in approximately equal read fractions. (D) Representative AAAAA/AAAAC calls within a homopolymeric region. Neighboring DRACH motifs are indicated in blue.

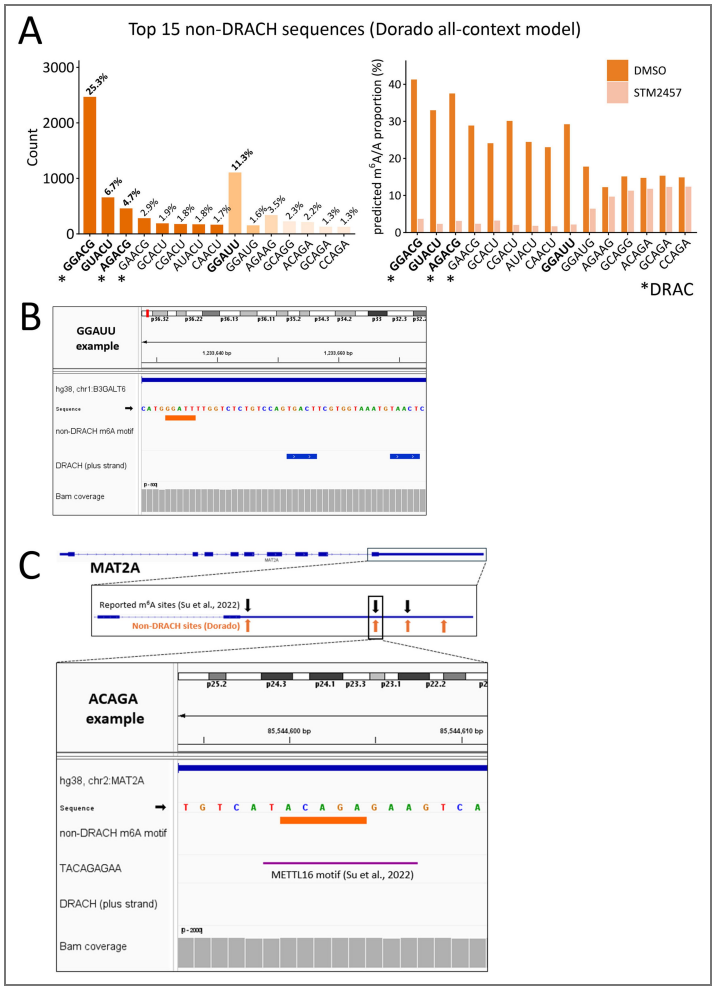


Figure S24. Non-DRACH m^{6}A calls predicted by the Dorado all-context model in NHDF data.

(A) Same as in Fig. S23A, but for m A sites predicted with the all-context Dorado model. (B) Representative locus showing an STM2457-sensitive GGAUU call without evidence of positional shift, sequence variation, or homopolymer context. (C) STM2457-insensitive ACAGA calls within *MAT2A*, shown relative to the METTL16 consensus (UACAGAGAA).

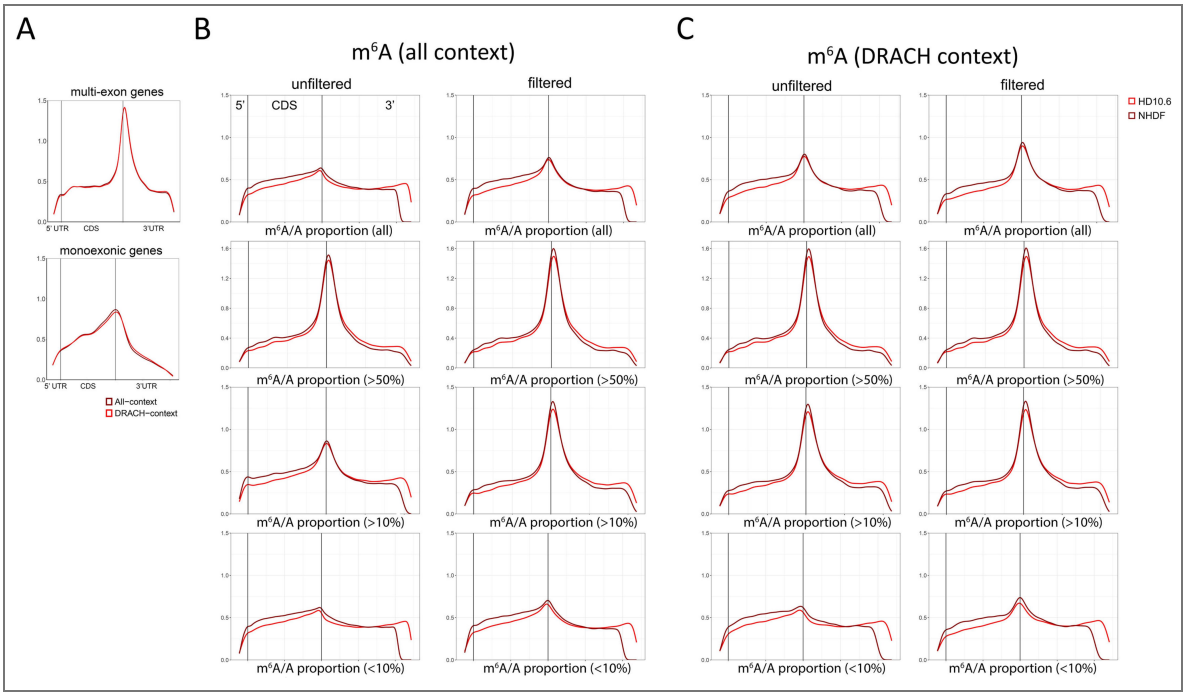


Figure S25. Metagene plots showing m A (all-context and DRACH-context) distribution profiles in NHDF and HD10.6 data filtered at different predicted m A/A proportions.

(A) Metagene plots showing the density of m^6A (all-context – dark red, DRACH-context – bright red) sites across all annotated poly(A) transcripts (upper plot) and monoexonic poly(A) transcripts (lower plot) from genome-aligned DMSO-treated NHDF datasets filtered with 0.98 modification probability, ≥ 20 coverage, and ≥ 10% predicted m^6A/A proportion. (B) The density of all-context m^6A sites (0.98 modification probability-unfiltered – left, filtered – right) is plotted across all annotated poly(A) transcripts of NHDF (dark red) and HD10.6 (bright red) datasets, filtered with different predicted m^6A/A proportions (descending: all proportions, ≥ 50%, ≥ 10% and < 10%. All datasets were filtered additionally for ≥ 20 coverage. (C) Same as (A), but for DRACH-context m^6A sites.

Data availability

Raw fastq and pod5 datasets generated in this study are available via the European Nucleotide Archive (ENA) under the accession number PRJEB101380. R scripts used in the analysis and visualization of data presented herein are available in the *https://github.com/DepledgeLab/HOLDEN* repository. The Snakemake workflow implementing the GLORI analysis pipeline is available in the *https://github.com/gp-micro/glori* repository.

Acknowledgements

We thank our colleagues in New York and Hannover for their continuous input on these studies and to Dr. Shino Murakami for access to an unpublished assay. DPD is supported by a German Centre for Infection Research (DZIF) Associate Professorship and the NIAID grants R01-AI170583 and R01-AI152543. DPD also receives funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2155, project number 390874280. EL is supported by the NIAID grant R01-AI170583 as well as the Hannover Biomedical Research School (HBRS) and the Center for Infection Biology (ZIB). DO is supported by the Walter Benjamin Programme of the DFG, project 554758329. ACW is supported by NIAID grants R01AI176335 and R01-AI170583.

Additional files

Supplemental Table S1.

Supplemental Table S2.

Supplemental Table S3.

Supplemental Table S4.

Additional information

Funding

Funder	Grant reference number
German Center for Infection Research	
National Institute of Allergy and Infectious Diseases	R01-AI170583
National Institute of Allergy and Infectious Diseases	R01-AI152543
National Institute of Allergy and Infectious Diseases	R01AI176335
Deutsche Forschungsgemeinschaft	390874280
Deutsche Forschungsgemeinschaft	554758329

Author ORCID iDs

Denise Ohnezeit: https://orcid.org/0000-0002-2346-2545
Adam T Silhavy: https://orcid.org/0009-0003-4308-9420
Jianheng Liu: https://orcid.org/0000-0003-0216-1951
Luke S Nicholson: https://orcid.org/0009-0005-8203-7249
Alejandro Pironti: https://orcid.org/0000-0002-7759-6776
Samie R Jaffrey: https://orcid.org/0000-0003-3615-6958
Daniel P Depledge: https://orcid.org/0000-0002-4292-0599
Angus C Wilson: https://orcid.org/0000-0002-5016-4164

References

Rand A (2024) Modkit. GitHub. version: v0.4.1 https://github.com/nanoporetech/modkit

Abebe JS, Verstraten R, Depledge DP (2022) Nanopore-Based Detection of Viral RNA Modifications. *mBio* **13**:e03702-03721 https://doi.org/10.1128/mbio.03702-21 | PubMed

Anreiter I, Mir Q, Simpson JT, Janga SC, Soller M (2021) New Twists in Detecting mRNA Modification Dynamics. *Trends Biotechnol* **39**:72-89 https://doi.org/10.1016/j.tibtech.2020.06.002 | PubMed

Cappannini A, Ray A, Purta E, Mukherjee S, Boccaletto P, Moafinejad SN, Lechner A, Barchet C, Klaholz Bruno P, Stefaniak F, *et al.* (2023) MODOMICS: a database of RNA modifications and related information. 2023 update. *Nucleic Acids Research* **52**:D239-D244 https://doi.org/10.1093/nar/gkad1083 | PubMed

Cruciani S, Delgado-Tejedor A, Pryszcz LP, Medina R, Llovera L, Novoa EM (2025) De novo basecalling of RNA modifications at single molecule and nucleotide resolution. *Genome Biology* **26**:38 https://doi.org/10.1186/s13059-025-03498-6 | PubMed

Cruciani S, Novoa EM (2026) The new era of single-molecule RNA modification detection through nanopore base-calling models. *Nat Rev Mol Cell Biol* **27**:10-18 https://doi.org/10.1038/s41580-025-00896-3 | PubMed

Diensthuber G, Milenkovic I, Llovera L, Milovanovic A, Pelizzari F, Novoa EM (2026) Systematic benchmarking of dorado basecalling models for RNA modification detection with highly multiplexed nanopore sequencing. *Nucleic Acids Research* **54** https://doi.org/10.1093/nar/gkag349 | PubMed

Diensthuber G, Novoa EM (2025) Charting the epitranscriptomic landscape across RNA biotypes using native RNA nanopore sequencing. *Mol Cell* **85**:276-289 https://doi.org/10.1016/j.molcel.2024.12.014 | PubMed

Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M, *et al.* (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**:201-206 https://doi.org/10.1038/nature11112 | PubMed

Garalde DR, Snell EA, Jachimowicz D, Sipos B, Lloyd JH, Bruce M, Pantic N, Admassu T, James P, Warland A, *et al.* (2018) Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* **15**:201-206 https://doi.org/10.1038/nmeth.4577 | PubMed

Ghohabi Esfahani N, Stein AJ, Akeson S, Tzadikario T, Powell C, Daneshvar Kakhaki P, Jain M (2026) Assessment of nanopore RNA modification calling in human cell lines and synthetic systems. *Genome Biology* **27**:190 https://doi.org/10.1186/s13059-026-04096-w | PubMed

Gilbert WV, Nachtergaele S (2023) mRNA Regulation by RNA Modifications. *Annu Rev Biochem* **92**:175-198 https://doi.org/10.1146/annurev-biochem-052521-035949 | PubMed

Gleeson J, Madugalle SU, Wan CY, McLean C, Bredy TW, De Paoli-Iseppi R, Clark MB (2025) Isoform-level profiling of m(6)A epitranscriptomic signatures in human brain. *Sci Adv* **11**:eadp0783 https://doi.org/10.1126/sciadv.adp0783 | PubMed

Guo W, Ren Z, Huang X, Liu J, Shao J, Ma X, Wei C, Cun Y, He J, Zhang J, *et al.* (2025) Single-molecule m6A detection empowered by endogenous labeling unveils complexities across RNA isoforms. *Molecular Cell* https://doi.org/10.1016/j.molcel.2025.01.014 | PubMed

He PC, He C (2021) . m(6) A RNA methylation: from mechanisms to therapeutic potential. *Embo j* **40**:e105977 https://doi.org/10.15252/embj.2020105977 | PubMed

He PC, Wei J, Dou X, Harada BT, Zhang Z, Ge R, Liu C, Zhang LS, Yu X, Wang S, *et al.* (2023) Exon architecture controls mRNA m(6)A suppression and gene expression. *Science* **379**:677-682 https://doi.org/10.1126/science.abj9090 | PubMed

Hendra C, Pratanwanich PN, Wan YK, Goh WSS, Thiery A, Göke J (2022) Detection of m6A from direct RNA sequencing using a multiple instance learning framework. *Nature Methods* **19**:1590-1598 https://doi.org/10.1038/s41592-022-01666-1 | PubMed

Hewel C, Wierczeiko A, Miedema J, Friedrich J, Hofmann F, Weißbach S, Dietrich V, Holthöfer L, Haug V, Mündnich S, *et al.* (2025) Direct RNA sequencing enables improved transcriptome assessment and tracking of RNA modifications for medical applications. *Nucleic Acids Research* **53**

https://doi.org/10.1093/nar/gkaf1314 | PubMed

Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**:44-57 https://doi.org/10.1038/nprot.2008.211 | PubMed

Jain M, Olsen HE, Paten B, Akeson M (2016) The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology* **17**:239 https://doi.org/10.1186/s13059-016-1103-0 | PubMed

Kang G, Hwang H, Jeon H, Choi H, Chang HR, Yeo N, Park J, Son N, Jeon E, Lim J, *et al.* (2025) Comprehensive discovery of m(6)A sites in the human transcriptome at single-molecule resolution. *Nat Commun* **17**:664 https://doi.org/10.1038/s41467-025-67417-w | PubMed

Killick R, Eckley IA (2014) . changepoint: An R Package for Changepoint Analysis. *Journal of Statistical Software* **58**:1-19 https://doi.org/10.18637/jss.v058.i03

Killick R, Fearnhead P, Eckley IA (2012) Optimal Detection of Changepoints With a Linear Computational Cost. *Journal of the American Statistical Association* **107**:1590-1598 https://doi.org/10.1080/01621459.2012.737745

Körtel N, Rücklé C, Zhou Y, Busch A, Hoch-Kraft P, Sutandy FXR, Haase J, Pradhan M, Musheev M, Ostareck D, *et al.* (2021) Deep and accurate detection of m6A RNA modifications using miCLIP2 and m6Aboost machine learning. *Nucleic Acids Res* **49**:e92 https://doi.org/10.1093/nar/gkab485 | PubMed

Köster J, Rahmann S (2012) Snakemake--a scalable bioinformatics workflow engine. *Bioinformatics* **28**:2520-2522 https://doi.org/10.1093/bioinformatics/bts480 | PubMed

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078-2079 https://doi.org/10.1093/bioinformatics/btp352 | PubMed

Linder B, Grozhik AV, Olarerin-George AO, Meydan C, Mason CE, Jaffrey SR (2015) Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nature Methods* **12**:767-772 https://doi.org/10.1038/nmeth.3453 | PubMed

Liu C, Sun H, Yi Y, Shen W, Li K, Xiao Y, Li F, Li Y, Hou Y, Lu B, *et al.* (2023) Absolute quantification of single-base m(6)A methylation in the mammalian transcriptome using GLORI. *Nat Biotechnol* **41**:355-366 https://doi.org/10.1038/s41587-022-01487-9 | PubMed

Liu JF, Hawley BR, Nicholson LS, Jaffrey SR (2025) Decoding m(6)Am by simultaneous transcription-start mapping and methylation quantification. *eLife* **13** https://doi.org/10.7554/elife.104139.3 | PubMed

Luo T, Xu M, Wang M, Chen F, Shi J (2026) Systematic evaluation of computational tools for multitype RNA modification detection using nanopore direct RNA sequencing. *Nat Methods* **23**:438-450 https://doi.org/10.1038/s41592-025-02974-y | PubMed

Luo Z, Ma Q, Sun S, Li N, Wang H, Ying Z, Ke S (2023) Exon-intron boundary inhibits m(6)A deposition, enabling m(6)A distribution hallmark, longer mRNA half-life and flexible protein coding. *Nat Commun* **14**:4172 https://doi.org/10.1038/s41467-023-39897-1 | PubMed

Meyer KD (2019) DART-seq: an antibody-free method for global m6A detection. *Nature Methods* **16**:1275-1280 https://doi.org/10.1038/s41592-019-0570-0 | PubMed

Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason Christopher E, Jaffrey Samie R (2012) Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTRs and near Stop Codons. *Cell* **149**:1635-1646 https://doi.org/10.1016/j.cell.2012.05.003 | PubMed

Moshitch-Moshkovitz S, Sevilla-Sharon M, Ashwal-Fluss R, Glick-Saar E, Rechavi G, Dominissini D (2024) mRNA m6A detection. *Nature Reviews Methods Primers* **4** https://doi.org/10.1038/s43586-024-00365-9

Mudge JM, Carbonell-Sala S, Diekhans M, Martinez JG, Hunt T, Jungreis I, Loveland JE, Arnan C, Barnes I, Bennett R, *et al.* (2025) GENCODE 2025: reference gene annotation for human and mouse. *Nucleic Acids Res* **53**:D966-d975 https://doi.org/10.1093/nar/gkae1078 | PubMed

Murakami S, Jaffrey SR (2022) Hidden codes in mRNA: Control of gene expression by m(6)A. *Mol Cell* **82**:2236-2251 https://doi.org/10.1016/j.molcel.2022.05.029 | PubMed

Raymon HK, Thode S, Zhou J, Friedman GC, Pardinas JR, Barrere C, Johnson RM, Sah DWY (1999) Immortalized Human Dorsal Root Ganglion Cells Differentiate into Neurons with Nociceptive Properties. *The Journal of Neuroscience* **19**:5420-5428 https://doi.org/10.1523/jneurosci.19-13-05420.1999 | PubMed

Sendinc E, Shi Y (2023) RNA m6A methylation across the transcriptome. *Mol Cell* **83**:428-441 https://doi.org/10.1016/j.molcel.2023.01.006 | PubMed

Shen W, Sun H, Liu C, Yi Y, Hou Y, Xiao Y, Hu Y, Lu B, Peng J, Wang J, *et al.* (2024) GLORI for absolute quantification of transcriptome-wide m6A at single-base resolution. *Nature Protocols* **19**:1252-1287 https://doi.org/10.1038/s41596-023-00937-1 | PubMed

Su C-H D D, Tarn W-Y (2018) Alternative Splicing in Neurogenesis and Brain Development. *Frontiers in Molecular Biosciences* **5**:2018 https://doi.org/10.3389/fmolb.2018.00012 | PubMed

Su R, Dong L, Li Y, Gao M, He PC, Liu W, Wei J, Zhao Z, Gao L, Han L, *et al.* (2022) METTL16 exerts an m(6)A-independent function to facilitate translation and tumorigenesis. *Nat Cell Biol* **24**:205-216 https://doi.org/10.1038/s41556-021-00835-2 | PubMed

Sun H, Li K, Liu C, Yi C (2023) Regulation and functions of non-m(6)A mRNA modifications. *Nat Rev Mol Cell Biol* **24**:714-731 https://doi.org/10.1038/s41580-023-00622-x | PubMed

Sun H, Lu B, Zhang Z, Xiao Y, Zhou Z, Xi L, Li Z, Jiang Z, Zhang J, Wang M, *et al.* (2025) Mild and ultrafast GLORI enables absolute quantification of m6A methylome from low-input samples. *Nature Methods* **22**:1226-1236 https://doi.org/10.1038/s41592-025-02680-9 | PubMed

Verstraten R, Cetraro P, Fitzpatrick AH, Alwie Y, Frommeyer YN, Loliashvili E, Stein SC, Häussler S, Ouwendijk WJD, Depledge DP (2025) Defining expansions and perturbations to the RNA polymerase III transcriptome and epitranscriptome by modified direct RNA nanopore sequencing. *bioRxiv* https://doi.org/10.1101/2025.03.07.641986 | PubMed

Viehweger A, Krautwurst S, Lamkiewicz K, Madhugiri R, Ziebuhr J, Hölzer M, Marz M (2019) Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. *Genome Res* **29**:1545-1554 https://doi.org/10.1101/gr.247064.118 | PubMed

Xiao YL, Liu S, Ge R, Wu Y, He C, Chen M, Tang W (2023) Transcriptome-wide profiling and quantification of N(6)-methyladenosine by enzyme-assisted adenosine deamination. *Nat Biotechnol* **41**:993-1003 https://doi.org/10.1038/s41587-022-01587-6 | PubMed

Xie YY, Zhong ZD, Chen HX, Ren ZH, Qiu YT, Lan YL, Wu F, Kong JW, Luo RJ, Zhang D, *et al.* (2025) Single-molecule direct RNA sequencing reveals the shaping of epitranscriptome across multiple species. *Nat Commun* **16**:5119 https://doi.org/10.1038/s41467-025-60447-4 | PubMed

Xu H, Kong L, Cheng J, Al Moussawi K, Chen X, Iqbal A, Wing PAC, Harris JM, Tsukuda S, Embarc-Buh A, *et al.* (2024) Absolute quantitative and base-resolution sequencing reveals comprehensive landscape of pseudouridine across the human transcriptome. *Nature Methods* **21**:2024-2033 https://doi.org/10.1038/s41592-024-02439-8 | PubMed

Yang L, Tang L, Min Q, Tian H, Li L, Zhao Y, Wu X, Li M, Du F, Chen Y, *et al.* (2024) Emerging role of RNA modification and long noncoding RNA interaction in cancer. *Cancer Gene Ther* **31**:816-830 https://doi.org/10.1038/s41417-024-00734-2 | PubMed

Yang X, Triboulet R, Liu Q, Sendinc E, Gregory RI (2022) Exon junction complex shapes the m(6)A epitranscriptome. *Nat Commun* **13**:7904 https://doi.org/10.1038/s41467-022-35643-1 | PubMed

Yankova E, Blackaby W, Albertella M, Rak J, De Braekeleer E, Tsagkogeorga G, Pilka ES, Aspris D, Leggate D, Hendrick AG, *et al.* (2021) Small-molecule inhibition of METTL3 as a strategy against myeloid leukaemia. *Nature* **593**:597-601 https://doi.org/10.1038/s41586-021-03536-w | PubMed

Yu J, She Y, Ji S-J (2021) m6A Modification in Mammalian Nervous System Development, Functions, Disorders, and Injuries. *Frontiers in Cell and Developmental Biology* **9**:2021 https://doi.org/10.3389/fcell.2021.679662 | PubMed

Zaccara S, Ries RJ, Jaffrey SR (2019) Reading, writing and erasing mRNA methylation. *Nature Reviews Molecular Cell Biology* **20**:608-624 https://doi.org/10.1038/s41580-019-0168-5 | PubMed

Zhou Y, Ćorović M, Hoch-Kraft P, Meiser N, Mesitov M, Körtel N, Back H, Naarmann-de Vries IS, Katti K, Obrdlík A, *et al.* (2024) m6A sites in the coding region trigger translation-dependent mRNA decay. *Mol Cell* **84**:4576-4593. https://doi.org/10.1016/j.molcel.2024.10.033 | PubMed

Zou Y, Ahsan MU, Chan J, Meng W, Gao S-J, Huang Y, Wang K (2025) A comparative evaluation of computational models for RNA modification detection using nanopore sequencing with RNA004 chemistry. *Briefings in Bioinformatics* **26** https://doi.org/10.1093/bib/bbaf404 | PubMed

Denise Ohnezeit, Elene, Daniel P.Depledge Loliashvili (2025) Establishing benchmarks for quan ta ve mapping of m6A by Nanopore Direct RNA Sequencing. European Nucleotide Archive. ID PRJEB101380 https://www.ebi.ac.uk/ena/browser/view/PRJEB101380

Peer reviews

Reviewer #1 (Public review):

The authors have substantially strengthened the manuscript in response to my comments. In particular, they have improved the treatment of non-canonical calls, added further illustrative examples at the level of individual RNA variants, repeated pathway enrichment using an appropriate expression-matched background, assessed treatment-associated cell viability, and moderated several biological claims. The revised manuscript is now most compelling as a technical benchmarking and best-practice resource for high-confidence prediction of a common RNA chemical modification from nanopore direct RNA sequencing.

The matched-input comparison of two methods, use of unmodified RNA controls, and inhibitor treatment are important strengths. Together, these experiments provide solid support for the central conclusion that careful filtering and appropriate controls can substantially reduce likely false-positive calls and yield a reproducible set of high-confidence predictions. The added analyses also provide useful illustrations of the capacity of long-read sequencing to assign predictions to individual RNA variants, although they do not establish broad superiority over the short-read comparison method.

A few claims remain insufficiently supported. First, the discussion of non-DRACH sites should cite Cancer mutations rewire the RNA methylation specificity of METTL3-METTL14 (Science Advances, DOI: 10.1126/sciadv.ads4750). In particular, Figure 3C of that study shows that the normal METTL3-METTL14 enzyme retains measurable activity on non-DRACH substrates; this provides a useful precedent for describing the inhibitor-sensitive GGAUU calls as plausible candidates rather than treating non-DRACH calls, especially those detected only by GLORI, as presumptively artefactual. The revised manuscript appropriately notes technical and biological explanations for non-canonical calls and identifies the GGAUU sites as candidates requiring further validation, but this additional context would improve interpretation.

Second, the comparison between Dorado and GLORI remains overinterpreted. The matched-input design and inhibitor perturbation are important strengths, but the current analysis does not establish that sites detected only by GLORI are false positives or that Dorado is a reference standard. This would require, for example, direct comparison of inhibitor responsiveness among GLORI-only and shared sites, while accounting for coverage, conversion efficiency, and other method-specific biases; therefore, phrases such as "GLORI exhibits reduced specificity," "largely spurious," and "validated by an orthogonal approach" are not fully supported by the data presented.

Similarly, the thresholding approach should not be described as "unbiased." The analysis is useful and empirically informed, but it relies on selected parameters, a manually specified penalty, and predetermined decision criteria. Describing the procedure as "data-driven" or "empirically calibrated against matched unmodified RNA controls" would more accurately convey both its value and its limitations.

The additional RNA-variant examples are helpful, but they demonstrate an advantage of long-read data rather than broadly proving that Dorado "outperforms" GLORI for RNA-variant-level quantification. A more accurate formulation would be: "Direct RNA sequencing provides RNA-variant-resolved predictions that are not readily assignable using short-read GLORI data, as illustrated at several loci."

Finally, the authors now acknowledge that the functional significance of cell-type differences remains unresolved, but the abstract still presents the results as evidence of "finely tuned" regulation, and the discussion links them to neuronal function. The data provide solid evidence for differences in predicted modification proportions between cell types, but the evidence is incomplete for claims about functional consequences; these statements should therefore be framed as context or hypotheses for future study rather than as inferences from the present experiments.

https://doi.org/10.7554/eLife.110672.2.sa2

Reviewer #2 (Public review):

The revised manuscript has improved substantially. The authors have addressed the main concerns raised during the previous round by introducing an IVT-based calibration framework, evaluating background calls across probability thresholds, examining basecalling mismatches, and replacing "stoichiometry" with the more appropriate term "predicted modification proportion." The expanded discussion of non-DRACH predictions is also useful and appropriately balanced. Overall, the study now provides a valuable practical framework for Dorado-based RNA modification analysis.

I have only a few remaining comments, all of which should be addressable through clarification and minor textual revisions.

(1) The comparison between NHDF and HD10.6 cells does not appear to include independent biological replicates. The observed differences in predicted m6A/A proportions should therefore be presented as differences between the two datasets rather than as general cell-type-specific effects. The interpretation is already cautious in the Discussion, but similar wording should be used in the Abstract, Results heading, and title if applicable.

(2) The direction of the comparison in Figure 3D should be clarified. The percentages represent the fraction of GLORI sites that is also detected by Dorado. Thus, the analysis shows that Dorado increasingly recovers high-proportion GLORI sites. It does not directly demonstrate that high-proportion Dorado predictions are increasingly validated by GLORI. The text should reflect this distinction, and the figure should clearly identify which circle represents each method.

(3) The ratio of calls detected in IVT relative to biological samples is a useful empirical measure of background, but it is not formally a false-positive rate because the true status and testability of every site are not known. A term such as "IVT-to-sample call ratio" or "empirical background-call ratio" would be more precise and would not alter the main conclusion regarding the value of stringent filtering.

(4) The description of the change-point framework still contains the placeholder "at least X-fold more signal." The actual enrichment threshold should be provided. The authors should

also ensure that a log2 ratio threshold of 1.5 is not described as 1.5-fold enrichment, since these values are not equivalent.

These are relatively minor issues. The authors have responded constructively to the previous reviews, and I do not consider further experimental work necessary.

https://doi.org/10.7554/eLife.110672.2.sa1

Author response:

The following is the authors' response to the original reviews.

We thank you all for the careful evaluation of our manuscript and for the numerous constructive comments and suggestions. As noted by the reviewers, the study directly compares two distinct approaches for detecting m^6A modifications using the same RNA input material. The inclusion of IVT and STM2457-treated RNA as negative controls, as well as the high-depth long-read datasets from two different cellular systems, enable robust analysis and interpretation of the data, providing a rigorous and reproducible analysis workflow and a valuable reference for the rapidly expanding community aiming to use Oxford Nanopore Technologies' (ONT) latest flow cell chemistry and Dorado base calling tool.

We have revised the manuscript according to this feedback and believe the changes have greatly improved the clarity and utility of the work. We have, where possible, addressed all the points raised and have further discussed or acknowledged those that involve broader technical considerations beyond the scope of the present study.

> ***Reviewer #1 (Public review):***
>
> *Summary:*
>
> *The authors set out to evaluate how accurately direct sequencing of RNA can identify and quantify several chemical modifications on RNA molecules, focusing primarily on m6A. A central goal of the work is to compare this approach with an independent chemical-based method (glyoxal and nitrite-mediated deamination of unmethylated adenosines (GLORI), using the same RNA samples, in order to assess reproducibility, false-positive signals, and sensitivity across a range of detection strategies. The authors further aim to demonstrate the biological utility of this approach by applying it to two human cell types, primary human fibroblasts and HD10.6 neurons. While the manuscript also reports detection of additional RNA modifications (pseudouridine and m5C), the depth of analysis and strength of controls are greatest for m6A, which forms the primary focus of the study.*
>
> *Strengths:*
>
> *A strength of this work is the direct comparison of two distinct measurement approaches performed on the same RNA input material; this has not been done in other recently published benchmarking studies evaluating the utility of the recent direct RNA sequencing for calling m6A. The authors systematically test multiple analysis models and show that, when appropriate filtering is applied, detection of modified sites is reproducible across software versions. The use of synthetic RNA standards and METTL3 inhibitors as negative controls helps to reinforce the overall results. The data show good agreement between the two methods at higher m6A modification levels, supporting the conclusion that direct RNA sequencing can reliably detect high-confidence modification sites. The authors also demonstrate that this approach can, in principle, provide information at the level of individual RNA variants (although only one example was provided), which is difficult to achieve with short-read methods. The methodology described here is likely to be useful to others seeking to apply similar approaches to*

> *identify and quantify m6A. The study also explores the detection of other RNA modifications, which highlights the broader potential of the approach, although these analyses are necessarily more exploratory given the more limited controls and data available.*
>
> *Weaknesses:*
>
> *Despite these strengths, several issues limit the interpretation of the results and should be clarified for readers.*
>
> *First, the authors appropriately address false-positive signals by estimating expected false-positive rates and by quantitatively comparing sequence motif enrichment before and after filtering. These analyses provide important support for the use of stoichiometry-based thresholds and demonstrate that filtering substantially improves specificity. However, even after filtering, a subset of detected sites remains outside the expected sequence context. It therefore remains unclear to what extent these non-canonical sites reflect genuine biology versus residual technical artifacts.*

The detected sites that were outside of the expected sequence context (0.1% of the sites per cell type) exhibited sensitivity to STM2457 treatment, supporting the conclusion that these sites represent genuine signals. This point has been clarified in the text (lines 424-455 and 563-578) and additional supporting data have been included in the supplementary figures (Supplementary Fig. S23 and S24).

> *Second, claims regarding the ability of direct RNA sequencing to resolve modification patterns across different RNA variants are supported by very limited evidence. The conclusion that this approach provides superior isoform-level quantification relative to short-read methods is based largely on a single gene example. While this case is interesting, it does not establish how widespread or general this advantage is. A broader analysis indicating how many genes show isoform-specific modification patterns detectable by this method, and how often these are missed by the comparison approach, would be necessary to support a general claim.*

We thank the reviewer for this valuable comment. To strengthen this point, we have added two additional examples illustrating isoform-specific m^6A modification patterns, as well as one monoexonic gene for comparison (new Supplementary Figure S22). *KRTAP1-5* is a monoexonic gene with a single detectable isoform and shows that Dorado detects more m^6A sites than GLORI. *HSPA8*, with five detectable isoforms, is a good example to show that two isoforms carry additional m^6A sites not resolvable by GLORI. *CITED2* has three detectable isoforms, and differential m^6A patterns can only be resolved with Dorado. We adjusted the text accordingly in lines 395-401.

> *Third, the biological interpretation of cell type-specific differences in modification levels remains underdeveloped. Although differences in modification stoichiometry are reported between fibroblasts and neuron-derived cells, the functional consequences of these differences are not addressed. It is unclear whether changes in modification levels are associated with differences in RNA abundance, stability, or translation. As a result, statements suggesting that these modifications fine-tune core cellular pathways are speculative and should either be supported with additional analyses or framed more cautiously.*

We have updated the text (lines 477-481) to ensure a more cautious interpretation of the observations.

> *Related to this point, differences in gene expression between the two cell types are a potential confounding factor. The pathway enrichment patterns presented appear biased toward particular functional categories, but without clear control for differential gene*

> *expression, it is difficult to determine whether the observed enrichment reflects cell type-specific regulation of RNA modification or simply differences in which genes are expressed. Clarifying how background gene sets were defined for these analyses would help readers interpret the results.*

We thank the reviewer for raising this important point. For the KEGG pathway enrichment analysis, we used genes harboring shared m^6A sites in both NHDF and HD10.6 cells as the input gene list, irrespective of their expression levels. We initially did not use a custom background gene set, which is also a valid point brought up by the reviewer. Hence, we repeated the analysis and used all genes expressed in both cell types (~13,000 genes) as background (i.e. excluding those only expressed in one cell type). We have revised the results (lines 467-468) and methods section (lines 740-749) accordingly. We replaced the plot in Figure 4H with the new plot.

> *The manuscript also suggests broader differences in overall modification levels between cell types, but this is not validated using an independent global assay. An orthogonal measurement of total modification levels on polyadenylated RNA (for example, dot blot) would help place site-specific stoichiometry differences in a clearer biological context.*

We tried to assess global changes in modification levels using a dot blot assay. However, even when loading 2 µg of RNA per sample, we did not detect visible differences between DMSO- and STM2457-treated samples in either cell type, which we expected to serve as a positive control. This most likely indicates that the dot blot assay is much less sensitive than our DRS-based sequencing approach. However, as we agree that for making a statement about broader differences between cell types, orthogonal validation would be necessary. Accordingly, we have adjusted our statements in lines 44-46 and 477-481.

> *Finally, the effects of the METTL3 inhibitor on these cell types are not fully characterized. While changes in m6A modification patterns are reported following treatment, the manuscript does not address whether the treatment affects cell growth or viability.*

Cell viability following STM2457 treatment was determined in NHDF and HD10.6 cells using the CellTiter-Glo® assay. HD10.6 cells are terminally differentiated and post-mitotic; therefore, proliferation/growth is not applicable to this cell type, and cell viability (assessed by CTG assay) is the relevant endpoint. We added the results, which showed no significant differences in either cell type, as a supplementary figure (Fig. S1). While we ensured that NHDFs showed normal growth behaviour when treating them with STM2457, because HD10.6 cells are non-proliferative, we assessed cellular health in each of the experiments conducted. We updated a description of the assay in the methods section, lines 637-643, and in the results section (lines 137-138).

> *Appraisal of conclusions and impact:*
>
> *Overall, the study provides an informative technical assessment of direct RNA sequencing for modification detection and establishes clear conditions under which the method performs well. The evidence strongly supports conclusions related to technical benchmarking, reproducibility, and the importance of filtering and controls, particularly for m6A. In contrast, conclusions regarding isoform-specific regulation and cell type-specific biological roles of RNA modification are less well supported by the data currently presented, and would benefit from either additional analysis or more restrained interpretation.*

> *The work is likely to have a meaningful impact as a practical reference for researchers using direct RNA sequencing, particularly by clarifying sources of false positives and the value of appropriate controls. With clearer limits placed on biological interpretation or more data presented in support of the biological interpretation, the study would serve as a valuable reference for users seeking to apply these technologies reliably.*

We thank the reviewer for this positive evaluation of our work. To address major concerns, we have adjusted our interpretation of the biological findings throughout the manuscript to better reflect the scope of the presented data. We believe these revisions more clearly distinguish the technical benchmarking aspects of the study from the biological insights that need further investigation.

> ***Reviewer #2 (Public review):***
>
> *Summary:*
>
> *In this study, the authors aim to establish a calibrated framework for detecting RNA modifications using long-read sequencing and apply it to compare modification patterns between fibroblasts and neuron-like cells. The work combines long-read sequencing, in vitro transcribed controls, methyltransferase inhibition, and comparison to an orthogonal sequencing-based method in an attempt to derive filtering strategies that reduce false positive modification calls. The authors further apply this framework to explore differences in modification levels between the two cell types. The resulting dataset may be of interest to researchers working on RNA modification detection using long-read sequencing technologies. Independent datasets across additional cellular systems can be useful for benchmarking computational methods and evaluating the behavior of modification detection models. However, the conceptual advance of the analytical framework presented here remains somewhat unclear, as many aspects of the analysis closely resemble strategies that have already been described in recent benchmarking studies.*
>
> *Strengths:*
>
> *A clear strength of the study is the generation of a relatively large long-read sequencing dataset together with several useful experimental controls, including in vitro transcribed RNA and pharmacological inhibition of the methyltransferase enzyme responsible for installing this modification. These controls are helpful for illustrating the challenges associated with distinguishing high-confidence modification sites from background signals. The inclusion of two different human cellular systems also provides an additional dataset that may be useful for benchmarking and cross-validation in the field. The study addresses a practically relevant question for the community, namely, how to reduce false positive calls in long-read sequencing-based RNA modification analyses.*
>
> *Weaknesses:*
>
> *The main weakness of the manuscript is its limited methodological novelty. Much of the analytical framework presented here closely follows benchmarking strategies that have already been described in recent studies of RNA modification detection using long-read sequencing. Several previous studies have evaluated modification-aware basecalling approaches, discussed the need for stringent filtering strategies, and compared long-read sequencing-based predictions with orthogonal mapping approaches. The manuscript would therefore benefit from a deeper engagement with the recent benchmarking literature and a clearer explanation of what conceptual or methodological advance the present study provides beyond these earlier analyses.*

We thank the reviewer for this constructive comment. At the time our study was conducted, relatively few benchmarking studies addressing these methodological aspects were available, and the analytical approach and thresholds applied in our study were therefore established independently of these reports. Several relevant studies (some initially as preprints) subsequently emerged within a relatively short period. Now that these studies are available, we have considered their findings, including the thresholds and filtering criteria applied, and have incorporated the relevant comparisons and methodological considerations into the revised manuscript. We have rewritten the introduction section (lines 99-117) and the discussion (lines 537-546, 547-549, 553-562) to incorporate newer studies and correctly integrate our study into the field. Furthermore, we consider these findings not to be supportive of our work, as they strengthen the overall interpretations and guidelines for ONT DRS users.

> *A second concern relates to the filtering strategy that forms the core of the proposed workflow. The manuscript applies several thresholds, including modification probability, stoichiometry, and read coverage cutoffs, but it is not clearly explained how these thresholds were determined. It remains unclear whether these cutoffs were derived from statistical calibration, empirical optimization using the presented dataset, or adopted from previous studies. Because the downstream conclusions depend strongly on these filtering choices, a clearer methodological justification would strengthen the work and help readers assess the robustness of the proposed framework.*

We thank the reviewer for raising this important point. For the modification probability thresholds, we compared the distribution of modification probabilities in the biological samples with that observed in the IVT controls. In the revised manuscript, we have integrated a new statistical framework that enables an unbiased determination of modification probability cutoffs for each dataset (see lines 174184, Methods, as well as Fig 2 and Supplementary Figs 5, 8, 9 and 10). We further report error rates relative to IVT across different modification probability cutoffs and probability ranges (lines 186-200). The validread coverage threshold was introduced as an additional safeguard against sites supported predominantly by reads that were neither confidently called as the canonical nucleotide nor confidently assigned as modified (since valid coverage includes only reads supporting the canonical base or the corresponding modification above the selected probability threshold (≥0.98 or ≥0.99, depending on the modification) (lines 233-240). Finally, the stoichiometry cutoff was also informed by comparison with the IVT controls. Although stringent modification-probability filtering substantially reduced background, the majority of residual sites detected in the IVT controls occurred at low modification proportions, particularly below 10%. We therefore introduced the stoichiometry threshold to further distinguish biological modification signal from low-level background calls observed in the unmodified IVT reference (lines 241-249).

> *The interpretation of the comparison between the two modification detection approaches also appears somewhat overstated. Differences between the methods are frequently interpreted as evidence that one approach produces large numbers of false positive calls, but the analyses presented do not fully exclude alternative explanations such as differences in sensitivity, sequencing depth, or methodological biases. A more cautious interpretation of these discrepancies would therefore be appropriate.*
>
> *Some discussion points also appear speculative. In particular, certain interpretations propose mechanistic explanations without presenting analyses that would allow these possibilities to be distinguished. Such interpretations would benefit from either additional supporting analyses or more cautious phrasing.*

We have revised the relevant sections accordingly to provide a more cautious interpretation of these findings (lines 583-592).

> *From a methodological perspective, the statistical robustness of the thresholds used throughout the analysis could also be discussed in more detail. Given the relatively modest read coverage cutoff applied in the study, low stoichiometry estimates may be strongly influenced by sampling noise, and fixed stoichiometry thresholds may therefore not correspond to a consistent level of confidence across sites. In addition, the manuscript relies heavily on fixed modification probability cutoffs to define high-confidence calls, but it does not discuss whether these scores are statistically calibrated or how they relate to expected error rates. Neural network outputs are often not well-calibrated probabilities, and interpreting these values as direct confidence estimates can therefore be problematic. Finally, modification detection models trained on known modification sites may capture sequence-context patterns present in the training data, meaning that motif enrichment or positional distributions along transcripts may partly reflect model biases rather than purely biological signals. A brief discussion of these limitations would help readers better interpret the robustness of the proposed filtering strategy and the downstream biological conclusions.*

We thank the reviewer for raising these important methodological considerations. We acknowledge that modification proportion estimates, particularly at sites with relatively low coverage and low modification proportions, may be influenced by sampling variability and may not accurately reflect the true modification proportion.

In terms of our filtering strategy, we performed false-positive rate analyses across multiple modification probability thresholds using repeated subsampling of biological and modification-free IVT datasets. In addition, we performed an independent change-point analysis to identify where modification signals in biological samples diverged from the IVT background. These analyses, together with the reduction in high-probability m^6A calls following METTL3 inhibition, provide complementary empirical support for the applied filtering criteria.

We also acknowledge that sequence-context features learned by modification-detection models may influence predicted motif distributions. We therefore performed additional analyses of non-DRACH predictions, identifying both technical sources of erroneous calls and non-canonical sites consistent with genuine biological modifications, including STM2457-sensitive sites.

> *Overall, while the dataset may be of interest to the community, the extent to which the study advances current methodological understanding beyond recent benchmarking efforts remains limited.*

We thank the reviewer for their comments and critiques. The revised version of the manuscript has addressed these points (as detailed above), and we consider the fact that other recently published benchmarking studies come to broadly similar conclusions provide additional support to our work.

Notably, however, our study remains the only one to use the same RNA as input for distinct sequencing approaches (i.e. nanopore vs. GLORI-Seq).

> *Minor comments:*
>
> *The discussion of the "DRACH" versus "all-context" outputs would benefit from greater technical precision. The statement that the number of sites within DRACH motifs identified by the all-context approach was nearly identical to the number reported by the DRACH model may suggest that these outputs derive from fundamentally different predictive models. As I understand it, the underlying neural network is the same, whereas the distinction lies primarily in the classification context. Clarifying this explicitly in the manuscript would improve interpretability and avoid potential confusion for readers.*

We have updated the text accordingly to ensure greater technical precision (lines 342-350).

> *The manuscript compares results obtained with different basecalling and modification settings but refers primarily to Dorado software versions. This may be misleading, as software version and model version are not necessarily equivalent. Different basecalling or modification models can be used with the same software release, and newer software versions may still use older models. For clarity and reproducibility, the authors should report the exact basecalling and modification model names used in the analyses rather than referring only to the Dorado software version.*

We thank the reviewer for highlighting this distinction. We have revised the manuscript and figures accordingly and clarified the specific Dorado software version as well as the individual models used for each analysis (lines 315-330).

> ***Reviewer #2 (Recommendations for the authors):***
>
> *(1) To strengthen the methodological robustness of the proposed framework, the authors should clarify how the filtering thresholds used throughout the analysis were determined. If these thresholds were derived empirically from the present dataset, the authors could consider providing additional analyses demonstrating their robustness (for example, by showing how key results change across different probability or stoichiometry cutoffs). If the thresholds were adopted from previous studies, the relevant sources and the rationale for using them should be explicitly cited.*

We thank the reviewer for this recommendation. As outlined before, the filtering thresholds used in the analysis were empirically informed by comparison of the biological samples with the unmodified IVT controls. In the revised manuscript, we have implemented a statistical framework that enables the identification of appropriate modification probability cutoffs on a per-sample basis. Specifically, we report error rates relative to the IVT controls across different modification probability cutoffs and probability ranges (lines 186-200), together with statistically supported identification of modification probability cutoffs for each biological sample (by comparing with the corresponding IVT reference, lines 220-230). These analyses support the modification probability thresholds applied in the downstream analyses.

We have also clarified the rationale for the additional filtering criteria, such as valid coverage and stoichiometry cutoffs. (lines 233-240 and 241-249).

> *(2) Because the analysis relies strongly on modification probability scores, the authors may also wish to discuss how these scores relate to expected error rates. If possible, providing an analysis of how false positive rates behave across probability thresholds using the IVT controls included in the study could help readers better interpret the meaning of the chosen cutoff values.*

We thank the reviewer for this suggestion. As described above, we have now directly assessed how false-positive rates vary across modification probability thresholds using the unmodified

IVT controls. In the revised manuscript, we report error rates relative to the IVT controls across a range of modification probability cutoffs and within discrete probability intervals (lines 203–213). These analyses demonstrate that false-positive rates decrease with increasing modification probability and provide an empirical basis for interpreting the selected probability thresholds.

> *(3) For the comparison between the two modification detection approaches, the authors may consider providing additional analyses that clarify the source of discrepancies between the methods. For example, examining whether differences depend on read coverage, sequence context, or site stoichiometry could help readers better understand the conditions under which the two approaches converge or diverge.*

Thank you for this suggestion. We agree that analyses of read coverage, sequence context, and site stoichiometry could help explain the differences between the two methods. However, our current results suggest that the main reason for the discrepancy is the limited conversion efficiency of the GLORI dataset (see Fig. 3D). To properly assess the effects of coverage or sequence context, we would need a higher-quality GLORI dataset with better conversion efficiency. Generating such a dataset and repeating the analysis was beyond the scope of this study. We have therefore revised the text to make this limitation clearer and adjusted our interpretation accordingly (lines 343-361 and 583-593).

> *(4) The interpretation of modification patterns between fibroblasts and neuron-like cells could also be strengthened by additional analyses that account for potential confounding factors such as differences in transcript abundance, transcript coverage, or gene expression levels between the two datasets.*

We thank the reviewer for this important suggestion. Our comparison was restricted to m^6A sites shared between both cell types, minimizing potential biases from cell type-specific transcript detection. We agree that transcript abundance, coverage, and the effects of m^6A on transcript stability and translation should be considered to assess the biological significance of these differences, but such analyses are beyond the scope of this study. We have therefore rephrased our statements more cautiously (lines 478-482).

> *(5) In the description of the modification detection models, the manuscript would benefit from clearer reporting of the exact basecalling and modification model names used in the analyses. Because software version and model version are not necessarily equivalent, specifying the model identifiers would improve clarity and reproducibility.*

As mentioned above, we have revised the manuscript and figures accordingly and clarified the specific basecalling and modification models used for each analysis (lines 315-330).

> *(6) The manuscript would also benefit from clarifying the relationship between the "DRACH" and "allcontext" outputs, particularly whether these represent distinct predictive models or the same model applied in different classification contexts.*

As mentioned above, we have revised the corresponding paragraph to avoid ambiguity regarding the models used in each analysis (lines 315-330).

> *(7) Finally, the manuscript would strongly benefit from a more detailed discussion of the relationship between the present analysis and previous benchmarking studies of long-read RNA modification detection, which would help readers understand how the current work relates to and extends existing efforts in the field.*

We thank the reviewer for this helpful suggestion. We have expanded the discussion to better place our findings in the context of new benchmarking studies (538-547, 554-563).

> ***Reviewer #3 (Public review):***

In this study, the authors aim to establish a calibrated framework for identifying RNA chemical marks from direct RNA sequencing data using a modification-aware basecalling workflow, with a particular focus on N6-methyladenosine. By combining native RNA sequencing with an unmodified control transcriptome, enzyme inhibition, comparison across multiple software versions, and orthogonal validation using an independent mapping approach, the authors seek to define a best-practice pipeline for reducing falsepositive calls and improving confidence in quantitative interpretation across cell types.

A major strength of the work is the rigor of the benchmarking strategy. In particular, the inclusion of an unmodified control transcriptome is both important and useful, and the study provides compelling evidence that this control remains necessary for robust interpretation, despite being omitted in many current workflows. The comparison across software versions and the matched analysis with an independent sequencing-based approach also substantially strengthen the evidence presented. The work therefore makes a valuable contribution to the community by offering a more stringent analytical framework that will likely be broadly useful to groups applying native RNA sequencing to study RNA chemical marks.

The evidence supporting the main conclusions is solid overall. The authors convincingly show that stringent filtering substantially reduces false-positive calls and improves agreement with orthogonal approaches, particularly at highly modified sites. The observation that many sites are conserved across cell types, while showing differences in relative modification levels, is also supported by the presented analyses.

At the same time, several conceptual issues limit the strength of some downstream interpretations. Most importantly, the manuscript repeatedly refers to the reported values as "stoichiometry," whereas the underlying software output is more appropriately interpreted as a statistical estimate of the proportion of aligned reads classified as modified. This distinction is important because the conclusions regarding celltype differences rely on quantitative comparisons of these values. In addition, the current calling framework depends on successful canonical base assignment before modification calling, which raises an important limitation: sites with the strongest signal deviations may be underrepresented if they are more likely to be miscalled during basecalling. This issue may be especially relevant for RNA marks that induce stronger mismatch signatures than N6-methyladenosine and should be more explicitly discussed. Overall, the authors largely achieve their primary aim of establishing a more rigorous and broadly applicable analytical framework for direct RNA sequencing-based modification detection. The work is likely to have a meaningful impact on the field, particularly by reinforcing the importance of appropriate negative controls and benchmarking standards. With clearer framing of the quantitative outputs and explicit discussion of current software limitations, this study will serve as a highly useful resource for the community.

Reviewer #3 (Recommendations for the authors):

I have several conceptual and methodological concerns.

(1) Please avoid the term "stoichiometry". A major concern is the repeated use of the term "stoichiometry." In the context of Dorado-based nanopore calling, this is fundamentally a statistical estimate of the proportion of aligned reads classified as modified, rather than a direct biochemical measurement of true molecular stoichiometry or occupancy. I strongly recommend replacing this term throughout the manuscript with:

- predicted m6A/A proportion

> *- relative modification proportionThis distinction is especially important because the manuscript draws biological conclusions from differences in these values across cell types. At present, the language risks overstating the quantitative certainty of the method.*

The reviewer is correct, and where applicable (text, figures), we have replaced the word "stoichiometry" with "predicted m^6A/A proportion".

> *(2) Canonical basecalling dependency requires explicit discussion.*
>
> *A key conceptual concern is that the modification caller requires the base to first be correctly called as A before assigning it as m6A. This point needs to be explicitly addressed in the manuscript. Because the proportion is being calculated only from reads that are first aligned and called as canonical A, I strongly encourage the authors to include a mismatch analysis and explicitly frame the reported values as proportions derived only from correctly aligned A-containing reads. This raises an important question: are the strongest signal-distorting sites being systematically excluded from analysis because they are more likely to be basecalled incorrectly? This is an important limitation of the ONT Dorado/Modkit framework and should be discussed directly.*

We thank the reviewer for raising this important point. We agree that the reported modification proportions do not capture all reads overlapping a given position. The proportions are calculated from reads in which the corresponding base is basecalled as an A and is confidently assigned as either canonical A or m^6A above the selected modification-probability threshold. Reads in which the position is basecalled as a different nucleotide, contain a deletion, are assigned another modification, or have an m^6A probability below the selected threshold are therefore not included in the proportion estimate. For the DRACH context model, reads containing a non-matching base at other positions within the required motif context are likewise excluded from modification calling. We have now clarified this limitation explicitly in the revised manuscript (lines 285-296). In addition, to assess the potential magnitude of this effect, we performed a mismatch analysis examining the percentage of reads in which the reference A was basecalled as another nucleotide and evaluated its relationship with the predicted m^6A/A proportion across sites in our filtered dataset (lines 297-313). This analysis provides information that is not captured by the predicted modification proportions.

> *(3) The concern above (2) is likely stronger for Pseudouridine and Inosine RNA editing.*
>
> *I think this issue is likely even more important for Pseudouridine and Inosine RNA editing than for m6A. For these modifications, the basecalling error at the modified site is expected to be substantially stronger than for m6A, which may leave very few aligned reads available for downstream proportion calculations. This could potentially explain why the authors require a higher modification probability threshold for these marks. I would strongly encourage the authors to explicitly discuss whether this basecalling dependency contributes to the need for the {greater than or equal to}0.99 threshold and whether this introduces a systematic detection bias.*

To investigate the potential impact of basecalling errors on modification proportion estimates, we examined alternative nucleotide calls at modification sites, including pseudouridine and inosine sites, focusing on positions where non-reference basecalls accounted for ≥25% of the reads overlapping the site (lines 297-305). At pseudouridine sites, the predominant alternative basecall was C, whereas at inosine sites, G was most frequently observed. These results indicate that modification-associated basecalling errors may lead to the exclusion of a fraction of reads from downstream modification proportion calculations. Consequently, the reported modification proportions may be underestimated at sites where the presence of a modification strongly affects canonical basecalling, introducing a potential detection bias.

Importantly, however, the modification probability thresholds used in our analysis were derived empirically by comparing modification probability distributions in the biological samples with those observed in the corresponding IVT and STM2457-treated controls. Basecalling accuracy was not considered in establishing these thresholds.

> *(4) Need better framing of multiple modifications on the same transcript.*
>
> *In the Introduction, I would encourage the authors to better position nanopore DRS relative to short-read approaches by explicitly noting that short-read methods are also limited in their ability to resolve multiple modifications on the same transcript molecule. This is an important conceptual advantage of native longread sequencing and deserves clearer emphasis early in the manuscript.*

We agree that the ability to examine the co-occurrence of multiple modifications on individual RNA molecules represents an important distinction between nanopore DRS and short-read approaches. We have therefore revised the Introduction to explicitly note the limitation (lines 79-80, 83-84).

> *(5) Clarify unclear experimental step.*
>
> *The stoichiometry distribution experiment (lines 189-195) in the Results section is difficult to follow.*

We thank the reviewer for pointing this out. We have revised and restructured this section to more clearly describe the stoichiometry analysis (lines 255-274).

> *(6) Discussion of non-DRACH sites.*
>
> *The manuscript interprets many non-DRACH GLORI sites as likely false positives. While this is a reasonable interpretation, I would like the authors to also discuss whether true cellular m6A may exist at non-DRACH motifs, and whether any non-canonical sites identified by nanopore DRS may represent biologically real events rather than technical artifacts.*

We agree and have added a new analysis (Fig. S23 and S24) addressing this point. Most non-DRACH calls derive from technical causes (SNPs, homopolymers, proximity to true DRACH sites), but a subset conforming to DRAC motifs, e.g., are STM2457-sensitive, indicating they likely represent genuine METTL3-dependent m^6A sites. We have added a description of the new figures in the results section 425456) and discuss this further in lines 564-579.

> *This is especially important given emerging evidence that non-canonical motifs can occasionally occur in vivo.*

We thank the reviewer and have added corresponding data and discussion, as pointed out above. While most all-context non-DRACH calls reflect similar technical artifacts, we identify

STM2457-sensitive GGAUU sites, e.g., and a distinct STM2457-insensitive, METTL16-like class (e.g., ACAGA in MAT2A), supporting that some non-canonical calls represent true m^6A sites.

> *Regarding lines 236-240: Are there also sites from nanopore sequencing that are unexpectedly positive?*

We thank the reviewer for this question and refer to our newly added figures (Figure S23 and S24) and explanations described above.

> *(7) Strongest point of the paper. I want to emphasize that one of the strongest contributions of this study is its clear demonstration that the modification-free transcriptome control is necessary. This should be stated even more explicitly as a best-practice recommendation, as the field increasingly trends toward analyses that omit such controls.*

We thank the reviewer for highlighting this important aspect of our study. We have strengthened this point in the Discussion (lines 610-611) and now more explicitly recommend using a modification-free transcriptome control as a best practice for reliable RNA modification detection by DRS.

> *Overall, I find this to be a strong manuscript, with its methodological rigor being a major strength. Addressing the concerns above would substantially improve both the clarity and interpretability of the work.*

We thank the reviewer for their positive assessment and constructive feedback. We believe that the revisions have improved the clarity of the manuscript and the interpretability of our findings.

https://doi.org/10.7554/eLife.110672.2.sa0