Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

RNA folding below the nanopore impairs direct RNA sequencing fidelity and modification calling

Marek Schwarz
Institute of Microbiology of the Czech Academy of Sciences
Josef Pánek
Institute of Microbiology, AS CR https://orcid.org/0000-0002-3388-3570
Saran Natarajan
Institute of Microbiology of the Czech Academy of Sciences
Veronika Kočárková
Institute of Microbiology of the Czech Academy of Sciences
Jana Wiedermannová
jana.wiedermannova@biomed.cas.cz
Institute of Microbiology of the Czech Academy of Sciences https://orcid.org/0000-0002-9979-0799

Article

Keywords:

Posted Date: July 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10315656/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

RNA folding below the nanopore impairs direct RNA sequencing fidelity and modification calling

Marek Schwarz[1], Josef Pánek[1], Saran Natarajan[1], Veronika Kočárková[1], Jana Wiedermannová[1*]
[1] Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic
[*]To whom the correspondence and material requests should be addressed.

RNA modifications are key for biological processes. Oxford Nanopore direct RNA sequencing (DRS) infers modifications from subtle ionic-current differences, but these signals are prone to misinterpretation due to systematic distortions. Here we identify an unrecognized source of DRS errors: RNA secondary structure folding below the nanopore during sequencing. Across engineered transcripts and modification-free bacterial and human transcriptomes, hairpins induce a reproducible upstream window with frequent sequencing errors, reduced base-quality, and false-positive calls across diverse RNA modifications. Disrupting hairpin formation by truncation or point mutation abolishes these effects, demonstrating a causal role for RNA folding. Signal-level analysis reveals localized perturbations in ionic current and dwell time upstream of secondary structures, consistent with altered translocation dynamics or interference at the nanopore exit. These findings define a general, structure-linked source of signal distortion, improve DRS data interpretation, and establish a basis for structure-aware model training and modification calling to enhance sequencing fidelity in basic research and translational applications.

Introduction

RNA is extensively decorated by chemical modifications—more than 170 have been described to date—and many influence RNA stability, translation and gene regulation across coding and non-coding RNA classes[1–4]. Mapping modification landscapes, particularly N6-methyladenosine (m6A), is widely used to study dynamic RNA regulation and has been proposed as clinically informative across diverse diseases[2,5–9]. To leverage RNA modification patterns as clinically useful biomarkers, detection methods must be straightforward, highly reliable and must capture dynamic changes in modification status. Oxford Nanopore (ONT) Direct RNA sequencing (DRS) is promising because it sequences full-length native RNA molecules while preserving single-molecule heterogeneity and, in principle, can detect RNA modifications directly from ionic-current signatures[10,11].

Despite improvements in chemistry, including an RNA-dedicated nanopore and faster motor protein[12,13], DRS still has lower per-read accuracy than short-read cDNA sequencing, with indels remaining prominent[14,15]. Besides basecalling errors, DRS modification calling also shows high false discovery rates[15]. Both sequencing and modification-calling errors are strongly context dependent[14–18], underscoring the need for continued methodological improvements, rigorous controls using non-modified RNA and orthogonal validation. DRS infers sequence and modification state from ionic current changes measured as RNA passes through the nanopore. Contemporary ONT basecallers use neural-network models to decode these signals into sequence, and modification callers interpret small residual, context-dependent deviations as evidence of chemical modifications[15,19,20]. As these signatures are subtle, specificity depends strongly on the quality of the primary signal and on its interpretation by basecalling and modification-calling models[19–21]. Consequently, systematic perturbations in ionic current or translocation dynamics can cause basecalling errors and false-positive modification calls, complicating de novo epitranscriptomic analyses[22–24]. The mechanistic causes of these distortions are only partially understood. Proposed contributors include low signal-to-noise ratio, chemical similarity of RNA nucleotides, and complex neighbor-dependent effects on the signal[21]. Distortions may also arise from sequence- and motor-interaction–dependent variability in translocation, for example in homopolymeric, C/U-rich or guanosine-rich pause sites[25].

RNA approaches the nanopore as a cDNA–RNA hybrid, which suppresses folding above the nanopore. However, once the RNA emerges below the pore, it may refold, potentially altering mechanical strand tension, ionic flow at the nanopore exit, or translocation kinetics in ways that could confound signal interpretation.

Here we demonstrate that RNA folding below the nanopore (in the trans compartment of the flow cell, Fig. 1d) is associated with DRS artefacts and false-positive modification calls. Using bacterial transcriptomes together with modification-free *in vitro*–transcribed (IVT) controls, we observe reproducible clusters of high-confidence modification calls and basecalling errors immediately upstream of stable RNA hairpins, including intrinsic transcription terminators and strong internal secondary structures. By disrupting hairpin formation through targeted truncations and mutations, we establish a causal link between RNA folding below the nanopore and loss of read fidelity. We further show localized deviations in ionic current and dwell time consistent with structure-linked perturbation of nanopore translocation dynamics. The generality was assessed by examining public datasets from additional organisms including a human cell line. Together, these findings reveal RNA folding in the trans compartment as a previously unrecognized source of DRS artefacts and motivate structure-aware interpretation of nanopore RNA sequencing data.

Results

In vivo and in vitro transcriptome of Bacillus subtilis reveal similar high-confidence modification patterns.

We used ONT's basecalling model Dorado v5.2.0 to detect RNA modifications in the Gram-positive model bacterium *Bacillus subtilis,* initially focusing on N6-methyladenosine (m^6A) calls across all sequence contexts. We revealed a striking accumulation of high-confidence modification calls (HCMs; modification probability ≥0.5) at the 3′ ends of coding sequences immediately upstream of intrinsic transcription terminators (Fig. 1a, bottom). Unexpectedly, the same pattern appeared in modification-free IVT controls (Fig. 1a, top), indicating that these HCMs represent false-positive modification calls. Basecalling errors, including mismatches, deletions and insertions, were also accumulated in these regions (Fig. 1b). To identify the cause of these DRS artefacts, we considered the biological role of the implicated RNA regions, bacterial terminators, which typically consist of a GC-rich hairpin followed by U-rich tract. During transcription, dynamic folding of this hairpin destabilizes the RNA–DNA hybrid within the transcription complex and promote its disassembly (Fig. 1c)[26,27].

We hypothesized that hairpin-folding could analogously affect RNA translocation through the nanopore and perturb the raw DRS signal, causing the enrichment of basecalling errors and false HCM calls upstream of the hairpins. When RNA emerges below the nanopore as a single strand, complementary sequences within a hairpin can refold. This can generate two outcomes: (i) changes in forces on the sequenced RNA molecule alter the kinetics of translocation through the nanopore and/or (ii) the folded RNA structure may partially obstruct the nanopore exit and alter ion flow, thereby changing the measured signal.

Disrupting downstream terminator hairpins eliminates false-positive HCM calls.

To test the causal effect of RNA terminator structures on basecalling, we selected four mRNAs that display false signals upstream of their terminators (*cspD, yfhH, iolW* and *yxkC*). We generated IVT transcripts in which the terminator hairpin was disrupted either by truncating the distal (3′) arm ("half", Extended Data Fig. 1) or by introducing point mutations that prevent hairpin base pairing ("mut"). In mutated/truncated transcripts, the false-positive HCM signal (m^6A, m^5C, inosine and pseudouridine) decreased to background levels (Fig. 1e-h), indicating that the downstream hairpin is required for the false-positive signal.

To test whether this effect depends on the proximity of the hairpin to the RNA 3′ end, we also prepared 3′-extended templates ("long"; Fig. 1e, Extended Data Fig. 1). In *cspD* long, false modification calls accumulated upstream of the hairpin, demonstrating that the artefact reflects the hairpin position rather than the transcript end.

Detailed per-nucleotide modification profiles for all constructs are shown in Extended Data Figs. 2-4, with summary tables provided for *cspD* (Fig. 2) and for *yfhH, yxkC* and *iolW* (Supplementary Tab. 1). Across all tested transcripts, multiple modification classes (m^6A, m^5C, inosine, 2′-O-mA, 2′-O-mU and 2′-O-mC) showed false positives within ~12 nt upstream of the 5′ end of the hairpin. These signals disappeared or were strongly reduced upon hairpin disruption. Newer Dorado models (v5.2.0 and later) reduced some of the structure-proximal artefacts compared to v5.0.0 (Extended Data Fig. 3e-j, Supplementary Tab. 1d,e), suggesting that secondary-structure-aware training could further improve DRS.

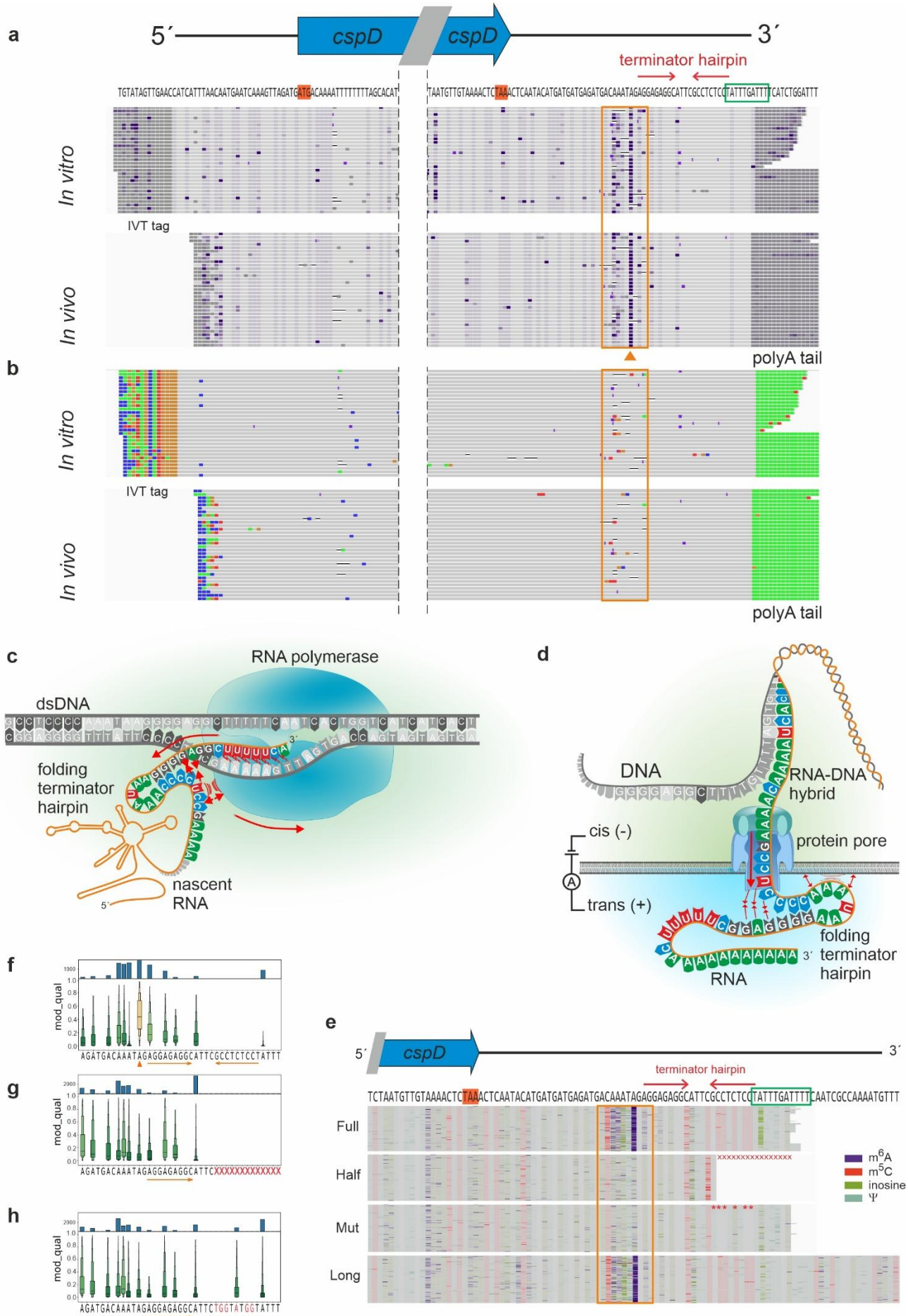

a
5′
cspD
cspD
3′
terminator hairpin
In vitro
IVT tag
In vivo
polyA tail
b
In vitro
IVT tag
In vivo
polyA tail
c
RNA polymerase
dsDNA
folding
terminator
hairpin
nascent
RNA
d
DNA
RNA-DNA
hybrid
cis (-)
protein pore
trans (+)
folding
terminator
hairpin
RNA
e
5′
cspD
3′
terminator hairpin
Full
Half
Mut
Long
m6A
m5C
inosine
Ψ
f
mod_qual
g
mod_qual
h
mod_qual

Fig. 1 | Regions enriched for false-positive modification calls and sequencing errors are located upstream of RNA secondary structures in ONT direct RNA sequencing. a, Direct RNA sequencing (DRS) reads mapped to the 3′ end of *cspD* from *B. subtilis* total RNA (bottom) and an unmodified IVT counterpart (top), basecalled with Dorado v5.2.0. Reads are visualized in IGV[28] and colored by m^{6}A probability (blue gradient). Gene model and nucleotide sequence are shown above; start/stop codons are indicated (orange), the intrinsic terminator hairpin is marked (red arrows) followed by the U-tract (green), and soft-clipped read ends are shown in dark gray. Grey triangles indicate positions with high-confidence modification (HCM; Dorado modification probability ≥ 0.5). **b**, Same region as in a, colored by basecalling errors (mismatches, insertions and deletions). Orange boxes mark regions enriched for sequencing errors and false modification signal; IVT-specific 5′ tag introduced by T7 transcription is indicated. **c**, Schematic of intrinsic transcription termination: nascent RNA hairpin formation destabilizes the transcription elongation complex. **d**, Model: RNA folding in the trans compartment (below the nanopore) perturbs ionic current and/or translocation kinetics, reducing basecalling fidelity and generating false modification signatures upstream of the folded structure. **e**, Disrupting terminator hairpin formation suppresses false-positive HCM calls. IGV views for *cspD* variants are colored by multiple modification types (m^{6}A, blue; m^{5}C, red; inosine, green; pseudouridine (Ψ), turquoise). Full, wild-type (genome-encoded) terminator; Half, truncation removing the distal (3′) stem arm; Mut, point mutations in the distal stem arm that prevent base pairing; Long, extended template producing a longer readthrough transcript that shifts the hairpin further upstream relative to the RNA 3′ end. **f–h**, Per-position m^{6}A calling for *cspD* full (**f**), half (**g**) and mut (**h**). Top: number of reads with non-zero m^{6}A modification called at each position (maximum depth 8,000). Bottom: distribution of m^{6}A call probabilities (mod_qual) represented by boxen plot where the center line denotes the median of the data. The terminator is indicated (orange arrows); mutated/deleted nucleotides are shown in red; the position with a high fraction of HCMs is marked (orange triangle).

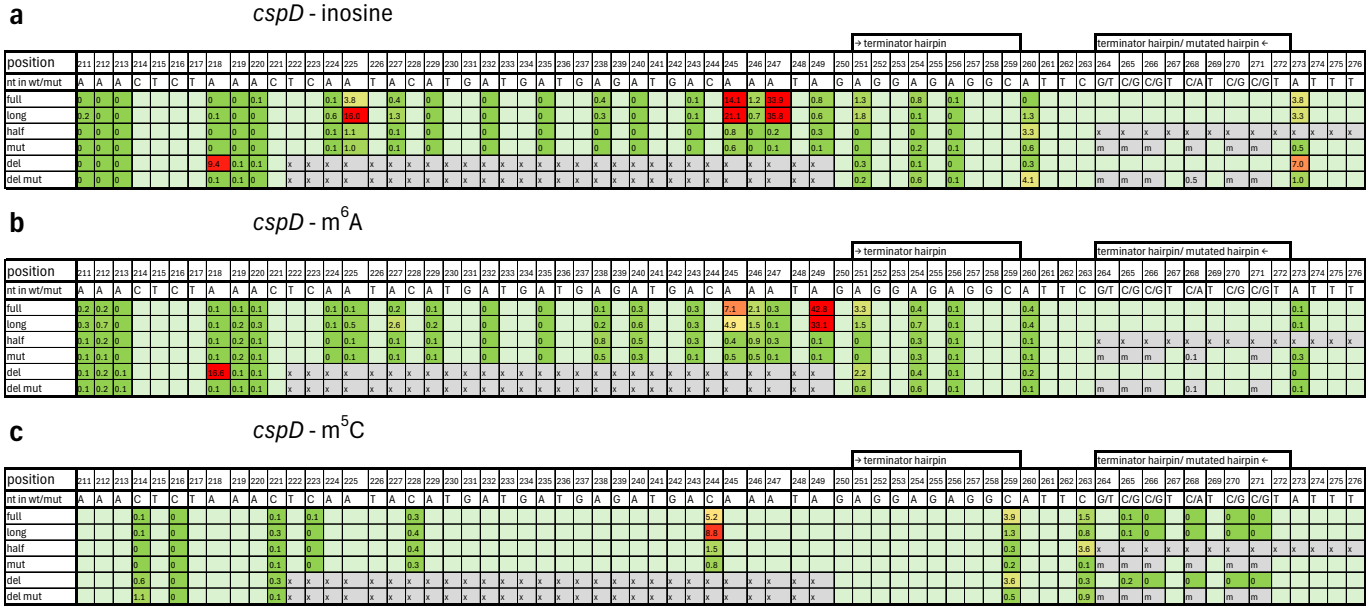


Fig. 2 | Dorado v5.2.0 modification calls in IVT *cspD* mRNA. Values indicate the percentage of reads called as modified at each position for (**a**) inosine, (**b**) m^{6}A and (**c**) m^{5}C, using default Modkit pileup filtering thresholds. Grey shading indicates mutations ("m") or deletions ("x") relative to Full; hairpin arms are marked in the header. Sites with >5% modified reads are color-scaled (yellow–red).

RNA secondary structure lowers upstream base quality scores (QS) and reduces DRS fidelity.

Next, we asked whether secondary structures affect QS of their upstream regions. We observed a localized decrease in QS at these positions. This decrease in QS persisted in "long" constructs, indicating that it does not depend on proximity to the RNA 3′ end. However, QS remained high when RNA folding was prevented (Fig. 3, Extended Data Fig. 4), demonstrating that it is caused by secondary structure formation. This pattern is similar to what we observed for false-positive HCMs (Fig. 1a,e) and increased error rates (Fig. 1b). Importantly, in *yfhH* and *iolW* transcripts,

hairpins caused HCMs without a corresponding decrease in QS (Extended Data Figs. 2,3). Thus, low QS itself is not sufficient indicator of false modification calls.

Hairpin structure can induce HCMs and low QS in different context

To evaluate the direct link between secondary structure, low QS and false-positive HCMs, we designed a transcript in which the *cspD* terminator hairpin is fused to a low-error region (high QS, no false HCMs; "del" construct, Extended Data Fig. 1). Consequently, the QS of the new hairpin-upstream region dropped dramatically (Fig. 3d), and new false HCMs—inosine and m^6A—appeared (Fig. 2a,b, position 218). No changes were observed in this region when fused to a mutated terminator with impaired base pairing (Fig. 2, lanes "del mut", Fig. 3e). We conclude that RNA hairpin is sufficient to perturb the nanopore signal and generate upstream sequencing artefacts.

Internal (gene-body) RNA hairpins reduce QS and induce false HCMs

Because terminators evolved specifically for their folding kinetics, we asked whether secondary structures within gene bodies also produce similar DRS distortions. We selected two sites in the *kdgT* coding region that exhibited a high frequency of sequencing errors in *in vitro* and *in vivo B. subtilis* transcriptomic datasets. In both cases we detected strong downstream secondary structures (MFE −14.9 and −14.79 kcal mol^{-1}, respectively), false HCMs, and drops in QS (~10–15 compared to ~30–40 of the flanking sequence; Fig. 4; Extended Data Fig. 5). When the secondary structure folding was prevented by introduced mutations, sequencing errors and HCM calls returned to background levels and QS increased. These results demonstrate that other secondary structures than terminators also affect the fidelity of nanopore sequencing and modification calling.

Low QS and false HCM colocalize with RNA secondary structures in other organisms

To assess the generality of the phenomenon, we analyzed publicly available DRS datasets from human cells[29] and *E. coli*[30] that include rigorous non-modified controls. Regions exhibiting false-positive HCM signal and/or elevated sequencing errors in both biological samples and IVT controls consistently localized upstream of stable RNA hairpins (Extended Data Fig. 6) indicating that the same structure-linked artefact occurs across organisms.

DRS signal features are affected by RNA secondary structure

The DRS raw signal is characterized by the observed ionic current (reflecting the identity of the RNA segment present in the pore) and dwell time (event length; reflecting the translocation dynamics). These signal features were analyzed to explain the artefacts observed upstream of RNA secondary structures. Across constructs containing secondary structures, we observed reproducible ionic current divergence from the Uncalled4-reported model predictions in a window immediately upstream of the structures. Event length of these regions also varied more strongly from the corresponding sequences with reduced folding potential (Fig. 5, Extended Data Fig. 7-8). The strongest effects were localized between ~15 nt upstream to ~5 nt within the predicted hairpin. Together, these analyses provide a mechanistic explanation for the reduced QS, showing that basecalling uncertainty arises from systematic perturbations in ionic current and translocation dynamics immediately upstream of secondary structures (Fig. 5i).

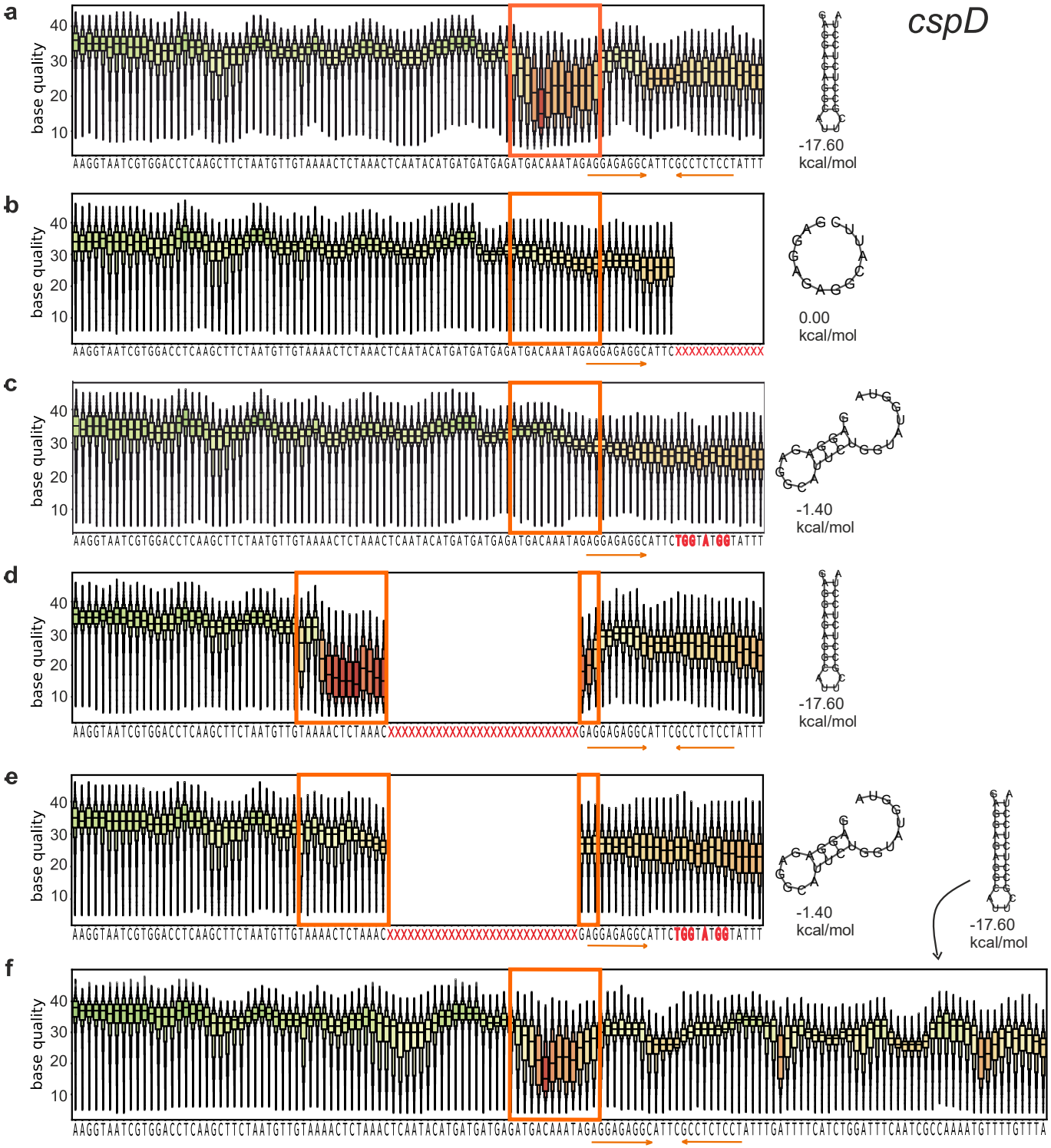


Figure 3 | RNA hairpins reduce upstream base quality scores (QS) in ONT direct RNA sequencing.
Panels **a–f** show distribution of the per-base quality score (QS) reported by the Dorado v5.2.0 model for nucleotides in the 3' terminal region of the *cspD* transcript encompassing the intrinsic terminator. Each position is colored by the median QS (red, lowest; green, highest). The distribution of QS at each base is represented by boxen plot where the center line denotes the median. The terminator hairpin is indicated by orange arrows; the predicted secondary structure of the terminator region (or mutated terminator region) and its minimum free energy (ΔG, kcal mol^{-1}) are shown. Mutated or deleted nucleotides are highlighted in red. **a,** *cspD* full. **b,** *cspD* half. **c,** *cspD* mut. **d,** *cspD* del. **e,** *cspD* del mut. **f,** *cspD* long.

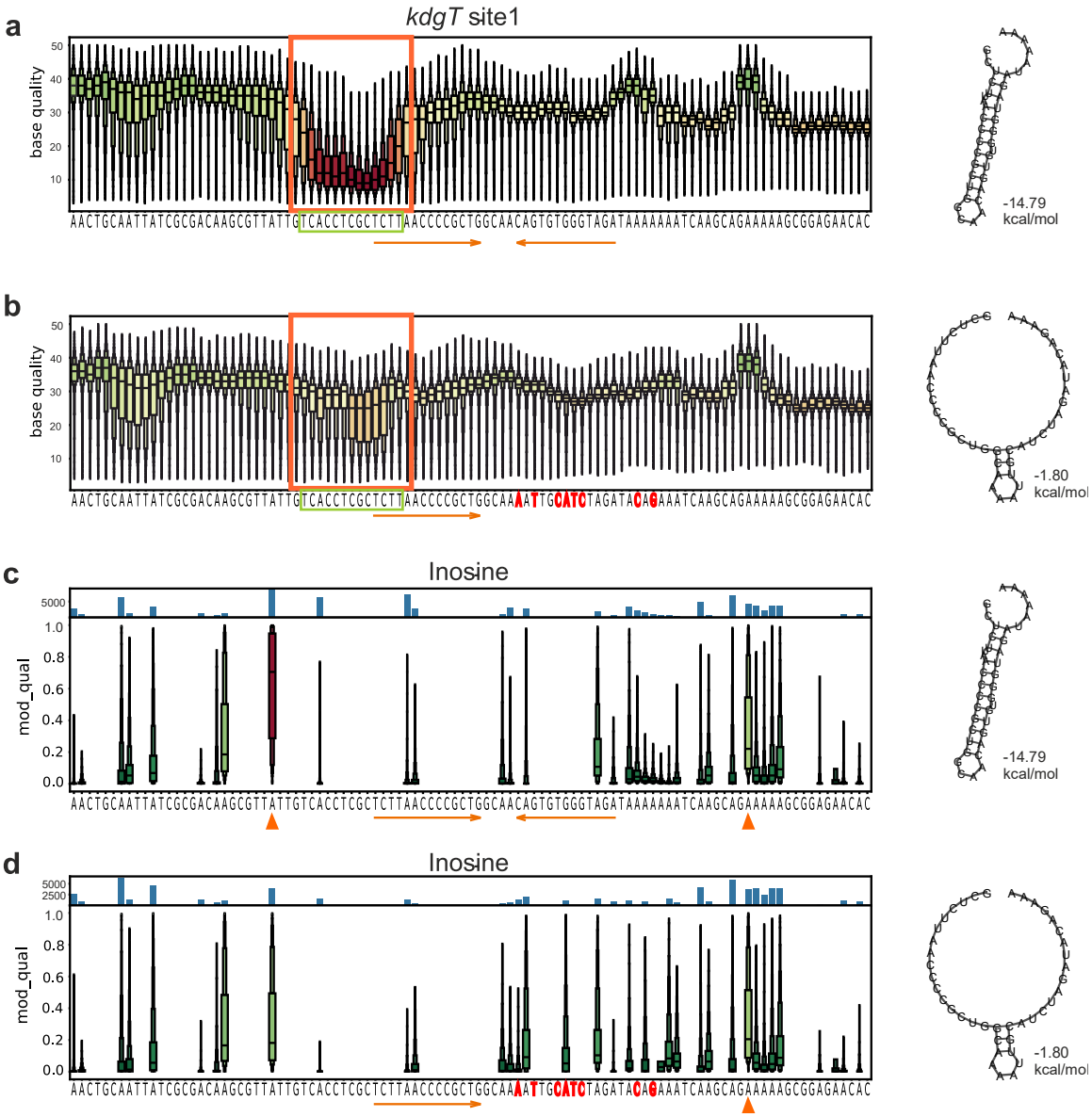


Figure 4 | Internal (gene-body) RNA hairpins reduce upstream QS and induce false modification calls. Panels **a**,**b** show boxenplots of per-base QS reported by Dorado v5.2.0 for nucleotides in the *kdgT* transcript surrounding a strong internal hairpin, displayed and annotated using the same conventions as in **Fig. 3**. Sequence encoding UC-rich RNA region is highlighted in a green box.

c,d, Per-position inosine calling for *wt* (**c**) and mut 1 (**d**) *kdgT* site1 region. Upper panel: number of reads with non-zero inosine modification called at each position (maximum depth 8,000). Lower panel: distribution of inosine call probabilities (mod_qual; Dorado/Modkit). The terminator is indicated (orange arrows); mutated/deleted nucleotides are shown in red; the position with a high fraction of HCMs is marked (orange triangle). **a,c**, *kdgT* wt; **b,d**, *kdgT* mut 1.

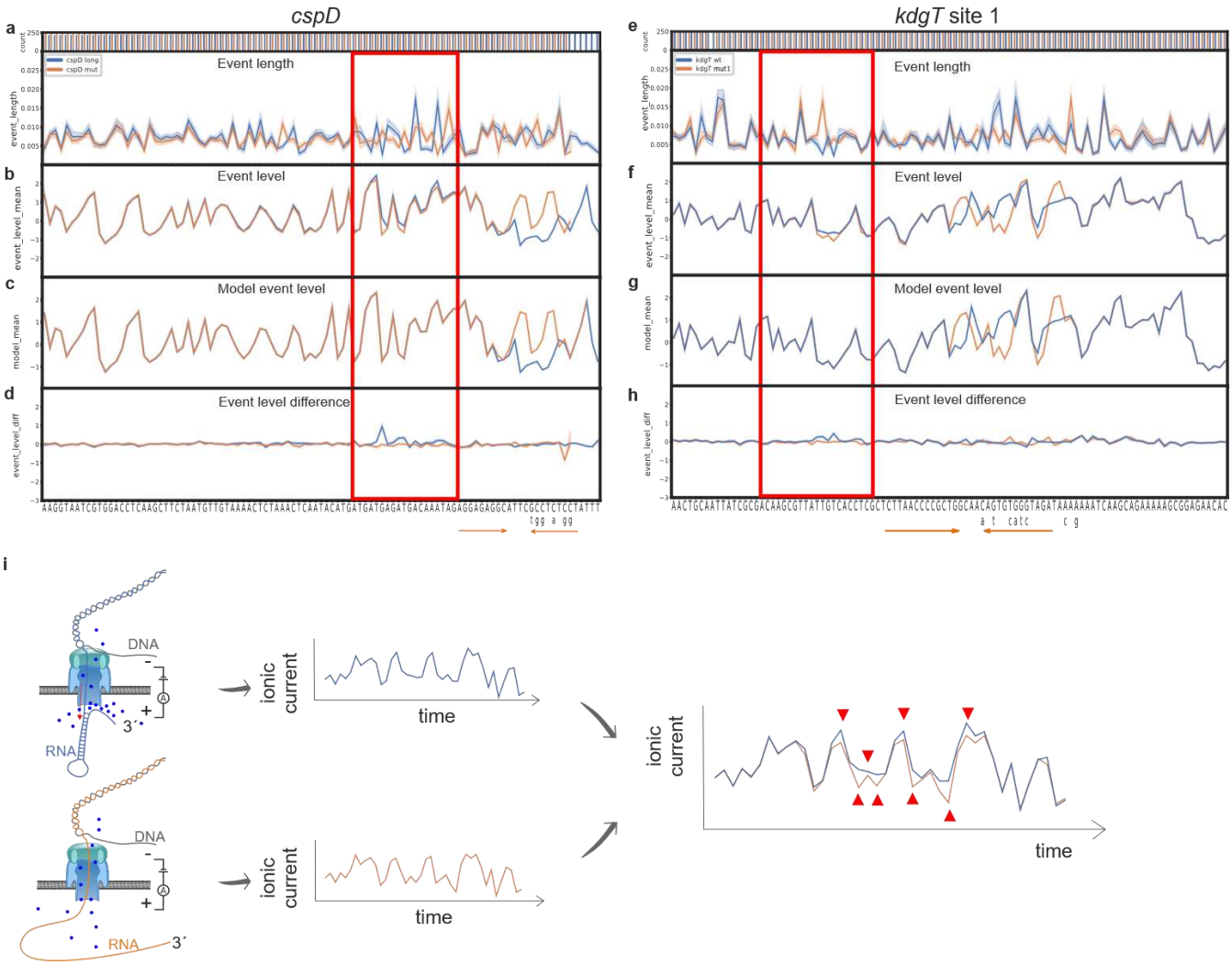


Figure 5 | Downstream hairpin formation perturbs dwell time and ionic current relative to model expectations.
Panels show position-wise event statistics obtained by signal/event alignment with Uncalled4 for *cspD* (**a—d**) and *kdgT* constructs (**e-h**) [long/wt (hairpin present; blue) and mut (hairpin base pairing disrupted; orange)] constructs. **a**,**e** Median of event length (dwell time) per position. **b**,**f** Median of observed event level (ionic current) per position. **c**,**g** Median of Uncalled4-reported model event level (model-predicted current) for the aligned sequence context. **d**,**h** Event level difference (observed minus model-predicted current means) per position. The terminator hairpin location is indicated beneath the plots (orange arrows), and the window showing the strongest structure-associated deviations immediately upstream of the hairpin is highlighted (red box). In all panels, a random subset of 250 reads was analyzed. Shaded bands indicate 95% confidence intervals; for **d,h** intervals correspond to a two-sided Z confidence interval for the difference of means ($\alpha = 0.05$). **i,** Proposed model of DRS signal perturbations by RNA secondary structures caused by ionic current disruption (ions depicted as blue circles) and altered RNA translocation dynamics (depicted as red arrow). Signal deviations are marked by red triangles.

Comparison of Dorado with other modification callers

As the artefacts arise from distortions in the underlying ionic current, we asked whether alternative signal-level modification callers are similarly compromised. We analyzed the IVT datasets with m6Anet[31] and SWARM[22], both trained primarily on eukaryotic and synthetic sequences.

m6Anet reported little to no m^{6}A signal in the IVT samples (Extended Data Fig. 9), consistent with the absence of true modifications. However, for bacterial samples this likely reflects the sequence-context dependence of the m6Anet build-in model, since it preferentially calls m^{6}A in eukaryote-specific DRACH motifs[31]. Whether this low false-positive rate compromises sensitivity could not be assessed because no validated m^{6}A sites have been identified in *B. subtilis.*

By contrast, SWARM calls false m^6A at similar hairpin-proximal positions as Dorado, and disrupting the downstream hairpin reduced false-positive calls for both tools (Extended Data Fig. 9). Together, these results indicate that distinct algorithm implementations can converge on the same structure-linked signal distortion, supporting a physical origin of the artefact—RNA folding in the trans compartment.

Low-QS regions are followed by increased secondary-structure potential transcriptome-wide

To extend our construct-based observations, we analyzed low-quality regions in the IVT *B. subtilis* transcriptome, intrinsically lacking any true modifications. We selected positions with a low per-base QS (Extended Data Fig. 10a) and analyzed −200 to +200 nt flanking regions for RNA secondary-structure potential and minimum free energy (MFE) (Fig. 6a,b). This analysis revealed a characteristic pattern: secondary-structure potential was locally reduced at the low-QS position itself but increased sharply downstream, while MFE showed the opposite trend indicating stable structures downstream of low QS regions. This organization mirrors our targeted construct experiments, in which low-QS regions occur immediately upstream of stable hairpins and are alleviated when hairpin formation is disrupted.

Motif analysis of low-QS regions identified an enriched UC-rich sequence (Fig. 6c,d). Together with prior reports that UC-rich sequences are associated with reduced nanopore basecalling accuracy[16] and our experimental evidence that downstream hairpins independently perturb signal quality, these results suggest a combined model in which UC-rich contexts and downstream secondary structures jointly contribute to reduced QS. Importantly, in our UC-rich construct (Fig. 4, green boxes) disrupting downstream hairpin base pairing markedly restored upstream QS, indicating that secondary structure contributes substantially to the artefact even in UC-rich contexts.

Transcriptome-wide colocalization of false-positive modification calls with low-QS regions and stable RNA hairpins

We analyzed the colocalization of false modification calls with (i) low-QS regions and (ii) secondary structures across the IVT transcriptome. In both cases, Fisher's exact test confirmed significant enrichment of modified positions in such regions (7.5-fold, $P = 7.5 \times 10^{-161}$ and 6.9-fold, $P = 1 \times 10^{-186}$, respectively). The colocalization of false modification calls was visualizied (Fig. 6g-i), showing the strongest ratio of false inosine and m^6A with low QS region centers and m^5C, m^6A and inosine upstream of secondary structures (Fig. 6h-i). Together, these transcriptome-scale analyses support our mechanistic model that hairpins are linked to reduced QS and elevated false modification calls in upstream sequence.

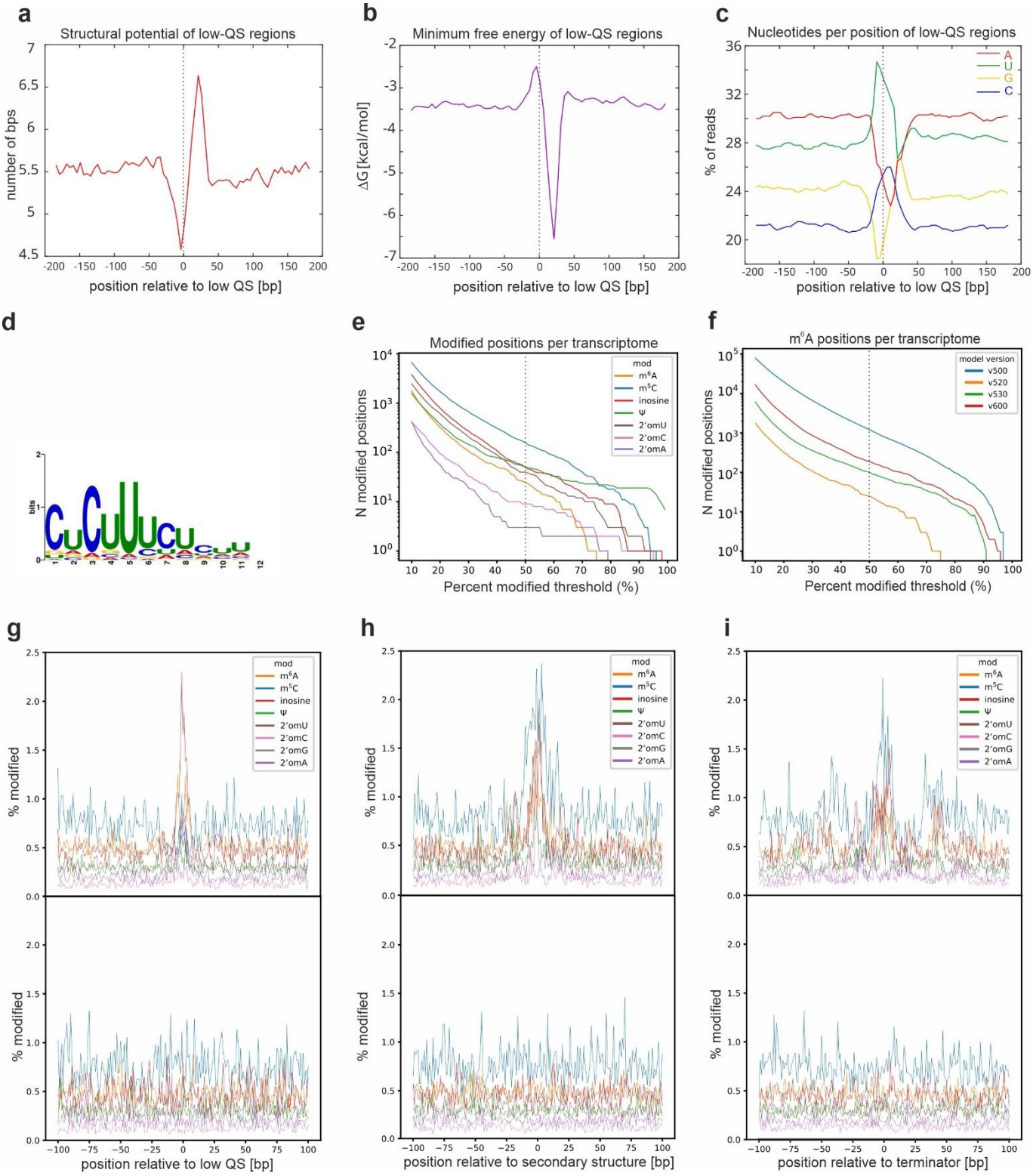


Figure 6 | *In vitro* transcriptome-wide analysis of low base quality, false modification calls and secondary structure. a, Secondary-structure potential around low-QS sites. Flanking sequences (−200 to +200 bp) of 1,946 positions with low per-base QS were analyzed for secondary structure. The plot shows number of base pairs within predicted secondary structure of the 30-nt sliding window (step 5 nt, x-axis indicates the first nucleotide of each window relative to the low-QS site). **b**, Mean minimum free energy (ΔG) of predicted RNA secondary structures across the same windows shown in **a**, indicating increased structural stability downstream of low-QS positions. **c**, Nucleotide composition around low-QS sites. Mean nucleotide frequencies (A, U, G and C) were computed across the same −200 to +200 bp regions around low-QS positions. **d**, UC-rich motif

enrichment in low-QS regions was identified within the −200 to +200 bp regions and was present in 66.9% of sequences (STREME[32]). **e**, *In vitro* transcriptome-wide distribution of false-positive modification calls: for each modification type (colors as indicated), the number of transcriptomic positions is plotted against the minimum fraction of reads required to be called as modified at that position (percent modified threshold). **f**, Model dependence of m^6A calls: as in **e**, but restricted to m^6A and showing different Dorado model versions (Table 1). For panels **e-f**, minimal coverage of 100 was required for each position. **g-i**, Transcriptome-wide colocalization of false-positive modification calls with low QS regions (**g**), predicted RNA secondary structures (**h**), and intrinsic transcription terminators (**i**) at the top panels, and randomized locations at bottom panels. Per-base modification percentages were computed with modkit localize, revealing increased ratio of false-positive modification sites (including m^6A, inosine and m^5C) within ±100 nt window.

Evolution of Dorado

During data collection and analysis, Dorado models were updated multiple times. We therefore analyzed IVT datasets by multiple model versions (Table 1) and found that, in some cases, the number of false-positive modification calls is even higher in newer models (Fig. 6f). These models still produce substantial false-positive signal in modification-free RNA, with tens of positions exceeding 50% modified reads and thousands exceeding 20% (Fig. 6e; Extended Data Fig. 10). The numbers also depend on coverage threshold: lowering it (for example, to 20× as used in some studies[11,33,34]) increased the number of positions with at least 50% modified reads to hundreds per bacterial transcriptome (Extended Data Fig. 10c). Together, these results indicate that model updates do not fully resolve the underlying artefact. Whether lower false-positive rates come at the cost of sensitivity will require benchmarking against orthogonally validated positives.

Discussion

Oxford Nanopore DRS profiles full-length native RNA molecule, enabling applications in basic research and clinical settings, including the use of RNA modifications as disease markers. However, analysis remains sensitive to unanticipated ionic-current distortions, therefore reproducible context-dependent error patterns persist, limiting its full potential[16]. Benchmarking studies emphasize that false modification discovery remain high and depends strongly on model version, thresholds and sequence context[11,29].

Here we identify RNA refolding below the nanopore as a previously underappreciated source of signal distortion, affecting translocation kinetics and current levels. Across transcriptomes and targeted IVT constructs, stable hairpins are associated with reduced basecalling fidelity and clusters of spurious HCM calls. Two non-mutually exclusive mechanisms could explain these effects: steric interference at the nanopore exit that alters ion flow and tension on the strand that perturbs translocation. The strongest QS-defects occur when a stable hairpin is preceded by a UC-rich segment within ~15 nt (Fig. 6a-d), consistent with the placement of these segments in the nanopore constriction and/or in contact with[35]. Although UC-rich contexts[16], our data show that downstream hairpins act as an independent driver of artefacts that is amplified in UC-rich context. Strong hairpins such as bacterial intrinsic terminators cause substantial signal disruption even in the absence of such context (Fig. 3, Extended Data Figs. 4,5) and UC-rich segments can display high QS, when not followed by secondary structures (Fig. 4b). This phenomenon is conceptually analogous to intrinsic transcription termination (Fig. 1c).

An earlier report described nanopore sequencing failure at long inverted duplicated DNA sequences (kilobase-scale) and proposed a tension-based explanation in which structure formation below the pore pulls the strand through faster, elevating current and reducing basecalling accuracy[36]. This mechanism aligns with our interpretation that trans-compartment folding can perturb motor kinetics and signal features in DRS. The convergence across nucleic acid types and vastly different length scales suggests that downstream folding might be general cause of artefacts in nanopore sequencing. Further investigation into possible effects of DNA structure formation might be of research interest.

Our findings have substantial implications for epitranscriptomic studies. Because the signal distortion caused by hairpin formation can match genuine modifications, *de novo* DRS modification calls are susceptible to false discoveries upstream of stable secondary structures. This vulnerability is not restricted to bacteria as RNA from

all organisms harbors strong RNA structures. Accordingly, candidate modification sites located upstream of strong hairpins should be interpreted cautiously unless supported by rigorous negative controls or orthogonal validation (for example, mass spectrometry–based measurements or antibody-based enrichment strategies)[37]. Differential isoform usage (e.g., alternative splicing) can further reposition secondary structures, potentially generating condition-dependent structure-linked artefacts that may not be captured by generic controls unless isoform-matched. Several benchmarking studies recommend prioritizing modification sites reported by multiple independent callers[34,38], however our findings indicate that this approach alone is insufficient when the underlying error originates in the primary signal.

The precise conditions under which these artefacts are observed—such as the minimum structure stability (ΔG), the relative contributions of stem length versus loop geometry, the tolerated spacing between the structured element and the pore, and whether more complex motifs (e.g., pseudoknots or multibranch folds) behave similarly—remain to be defined.

To address these artefacts in practice, we propose three complementary strategies:

(i) Modification-calling algorithms should incorporate a structure-aware quality-control step in which bases within a 15 nt window upstream of predicted strong hairpins are flagged or downweighted.

(ii) Future basecalling models should be explicitly trained using synthetic RNA constructs containing defined hairpins or using secondary-structure-aware data from non-modified IVT transcriptomes (including our dataset). This should help models to learn folding-induced signal perturbations across structural motifs encountered *in vivo*. Because ONT maintains the modified basecalling models, incorporating structure-aware features into these models is likely the most direct path to robust mitigation.

(iii) Suppressing or preventing hairpin formation in the DRS flowcell trans compartment provide the most reliable artefact mitigation strategy. These should include modifying the flow-cell trans-compartment buffer with chemical additives, denaturing agents or employing immobilized RNases. However, such interventions will require careful optimization.

In conclusion, trans-compartment RNA folding systematically compromises DRS fidelity and generates false-positive modification calls, underscoring the need for structure-aware mitigation strategies to ensure reliable nanopore-based epitranscriptomic measurements in both basic research and clinical diagnostics.

Methods

Oligonucleotides

Oligonucleotides used in this work are listed in Supplementary **Table 2**.

Bacterial growth and RNA isolation

B.subtilis (BSB1[39]; lab strain LK4269) was inoculated from the overnight culture and grown in LB medium at 37°C in Bioreactor RTS-8 (Biosan) at 2000 rpm. 2 ml aliquots were collected at the early stationary phase at OD_{600} ~2 by a short centrifugation (1 min) and flash freezing on dry ice and stored in -80°C until used. Each frozen cell pellet was resuspended in 240 µl TE buffer (10 mM TRIS-HCl pH 8.0, 1 mM EDTA), 60 µl LETS buffer (50 mM Tris–HCl pH 8.0, 500 mM LiCl, 50 mM EDTA pH 8.0, 5% SDS) and 600 µl acidic phenol/chloroform/isoamylalcohol (125:25:1, pH~4.5, Sigma-Aldrich). Lysates were sonicated (Bandelin sonopuls, VS 70 T probe, 3 × 40 s on ice with 1-min pauses between cycles, amplitude 50%), centrifuged (13500 rpm, 5 min, RT), the aqueous phase extracted two more times with acidic phenol/chloroform/isoamylalcohol and two times subsequently precipitated with 0.3 M Sodium Acetate and 3 volumes of ice-cold ethanol to ensure the proper removal of phenol. RNA was dissolved in Nuclease-free water and treated with DNase (TURBO DNA-free Kit, Ambion).

Direct RNA sequencing

20 µg of RNA was ribodepleted using *B. subtilis* (RNA-seq) riboPOOL (siTOOLs) kit. RNA yield was quantified by Qubit fluorometer (RNA HS Assay kit, Invitrogen), and the efficiency of rRNA depletion was verified by 2100 Bioanalyser (Agilent). The ribodepleted RNA was polyadenylated in 20 ml reaction containing 1 ml *E. coli* poly (A) polymerase

(New England Biolabs) in 1X Poly A buffer supplemented with 1 mM ATP for 1 min at 37°C. The reaction was stopped by the addition of 5 ml of 5 mM EDTA and cooled on ice. Polyadenylated RNA was purified by RNAClean XP beads (1.8 volume) and eluted in 9 ml of H_2O. RNA quantity and quality were analysed by Qubit and Bioanalyser, respectively.

The DNased, ribodepleted and polyadenylated RNA was used for DRS library preparation according to the ONT protocol without any modifications. The resulting library was loaded on PromethION RNA flowcell (R10.4.1) and sequenced on PromethION P2 solo device using default settings for DRS and live Super accurate (SUP) basecalling model.

Non-modified IVT library preparation

The DNased, ribodepleted and polyadenylated RNA was used as a template for non-modified library preparation. The protocol is based on the published protocol[37] with several modifications. Briefly: 3.5 µl of the sample was mixed with 1 µl of 100 µM oligo $d(T)_{23}VN$ primer (LK5600), 1 µl 10 mM dNTPs, 0.5 µl RNAsin Ribonuclease Inhibitor (Promega) and annealed by heating to 70°C for 5 minutes, followed by cooling on ice. In the next step, the RNA was reverse-transcribed into cDNA with the addition of T7 promoter on the 5´end by template switching as follows: annealing mixture from the previous step was supplemented with 2.5 µl of Template switching RT buffer (New England Biolabs), 1 µl Template switching RT enzyme mix (New England Biolabs), 1 µl of 100 µM TSO T7 oligonucleotide (LK5586, which includes the IVT recognition tag AGACTGCAGCATGGGG) and incubated 90 min at 42°C, 5 min at 85°C and cooled to 4°C. The template RNA was removed by RNase treatment with 1 µl of RNase H (New England Biolabs) in 1X RNAse H buffer. The resulting ssDNA was purified by RNAClean XP beads (0.8 volume) and eluted in 23 ml of H_2O.

Second-strand synthesis was performed from the whole sample by primer extension, adding 25 µl of Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs), 2 µl of 100 µM T7 oligonucleotide (LK5587) and incubating at 95°C for 1 min, 65°C for 10 min and cooled to 4°C. dsDNA was purified by RNAClean XP beads (0.8 volume), eluted in 9 ml of H_2O and quantified by Qubit (dsDNA HS Assay kit, Invitrogen).

The resulting dsDNA was used for *in vitro* transcription by MEGAscript T7 kit (Invitrogen) as follows. 2 µl of each nucleotide (ATP, GTP, CTP, UTP), 2 µl of 10X transcription buffer, 8 µl of template dsDNA, 2 µl of T7 RNA polymerase were incubated at 37°C for 3 hours in Thermocycler. The final IVT RNA library was DNAsed by TURBO DNA-free Kit (Ambion) and analyzed for quantity and quality by Qubit and Bioanalyser, respectively.

Wt and mutated IVT RNAs – DNA template synthesis

Templates for *in vitro* transcription were generated by PCR in 50 µl reactions using 200 ng genomic DNA (isolated from *B. subtilis* BSB1)[39] as a template and 0.5 µM primers (**Supplementary tab. 2**). The T7 promoter sequence, followed by IVT-recognition tag (AGACTGCAGCATGGGG) was introduced via the forward primers upstream of 5´ end of the coding sequence of the individual transcripts. Construct variants were produced by primer-encoded modifications as follows. Half constructs lacked the distal (3') arm of the intrinsic terminator hairpin and were generated by reverse-primer–introduced truncation. Long constructs were generated using a reverse primer positioned 15–40 bp downstream of the full-length design, thereby extending the template beyond the terminator. Because the terminator sequence remained present, reads corresponding to the extended (readthrough/run-off) RNA species were subsequently identified and selected bioinformatically (see "Sorting of individual IVT RNA species" below). Mut constructs were generated by introducing point mutations in the reverse primer to disrupt terminator base pairing and prevent hairpin formation. Del constructs were designed to reposition the terminator hairpin sequence downstream of a novel region (previously exhibiting high QS), by embedding the terminator hairpin sequence within the reverse primer. Del mut constructs were generated analogously, but with the distal-arm–disrupting mutations present in the terminator sequence within the reverse primer. All dsDNA PCR products were verified by agarose gel electrophoresis and gel-purified prior to transcription.

Chimeric PCR for mutated kdgT constructs

Chimeric PCR was used to generate templates for mutated *kdgT* constructs. In that case, two fragments for each construct (*kdgT* site1 and *kdgT* site 2) were prepared by PCR and purified from gel (using QIAquick gel extraction kit,

Qiagen). These fragments contained a 21-nt complementary region which was used to anneal the two parts [2.5 µl of each gel-purified fragment were combined in 10 µl final volume of 1X Phusion buffer, the mix was heated at 95°C for 2 min and slowly cooled (1°C min^{-1}) to 35°C]. 1 µl of the annealing mixture was used as a template for PCR with the outer primers (LK5768, LK5772), starting the PCR reaction by 3 min of elongation at 70°C and proceeding with standard PCR protocol. The resulting chimeric DNA was gel-purified to ensure the selection of the full-length constructs.

Wt and mutated IVT RNAs – T7 transcription

20 ng of each DNA template containing T7 promoter was used for *in vitro* transcription using MEGAscript T7 kit (Invitrogen) following manufacturer's instructions. Concentration of the resulting RNA was measured by Qubit and samples were (1.) pooled equimolarly in one tube, (2.) purified by Monarch RNA Cleanup kit (New England Biolabs), (3.) DNased by TURBO DNA-free Kit (Ambion), (4.) repurified by Monarch RNA Cleanup kit and (5.) polyadenylated by *E. coli* poly (A) polymerase (New England Biolabs) as described in section "Direct RNA sequencing". In parallel, the sizes of resulting RNA species were verified by radiolabeled *in vitro* transcription and polyacrylamide gel electrophoresis. The pooled RNA was sequenced following ONT DRS library preparation and sequencing protocols. Data for individual RNA species were sorted bioinformatically.

Sequencing of the strain B. subtilis BSB1 wt-derivative 168 trp+ (LK4269)

Strain LK4269 (BSB1[39]) was sequenced by Illumina and ONT. Genome was assembled with hybracter[40] pipeline using hybrid assembly approach and annotated with NCBI Prokaryotic Genome Annotation Pipeline (pgap)[41]. Sequencing data and assembled genome was deposited at NCBI under following accession numbers: BioProject PRJNA1173365, Accession CP171645.

Basecalling of the DRS

Raw DRS signal data were basecalled using Dorado (ONT; RRID:SCR_025883, https://github.com/nanoporetech/dorado/). Basecalling was performed with RNA004 SUP models that support modified-base inference, and reads were aligned to the *B. subtilis* reference genome (accession CP171645) using minimap2[42] with options -x map-ont.

Primary analyses were performed using the Dorado RNA004 SUP model v5.2.0. Additional basecalling model versions (v5.0.0, v5.3.0, and v6.0.0) were used only where explicitly indicated, for targeted comparisons and confirmatory analyses. The Dorado software releases, basecalling model versions, and enabled modified-bases are listed below; modification sets follow the RNA004 model compatibility documented in the Dorado model catalogue.

Table 1. Dorado models used in this work.

Dorado version	Model version	Modified bases	Referred to as
0.9.1	sup@v5.0.0	pseU m6A	v5.0.0
1.1.1	sup@v5.2.0	m5C_2OmeC inosine_m6A_2OmeA pseU_2OmeU 2OmeG	v5.2.0
1.4.0	sup@v5.3.0	m5C_2OmeC inosine_m6A_2OmeA pseU_2OmeU 2OmeG	v5.3.0
2.0.0	sup@v6.0.0	m5C_2OmeC inosine_m6A_2OmeA pseU_2OmeU 2OmeG	v6.0.0

Wt and mutated IVT RNAs - sorting of individual transcript variants

Because IVT constructs were sequenced as a two independent DRS libraries and pooled bioinformatically, basecalled reads were then separated into individual transcript variants (**Extended Data Fig. 1**). Only reads with a minimum mean read quality of 10 and mapping to defined regions of the *B. subtilis* CP171645 genome corresponding to *cspD, iolW, yfhH, yxkC* and *kdgT* were retained. Reads mapping to each gene were further classified into individual transcript variants (full, wt, mut, long, half, del, del mut, mut 1, mut 2) based on construct-specific sequence features defined in **Extended Data Fig. 1**. Genomic coordinates of the selected regions

and the regular expressions used for variant assignment are provided in Supplementary File: [SF terminator sequences 4]. Reads assigned to each variant were subsequently realigned to the corresponding transcript reference using minimap2[42].

Pileup generation and plotting

Base quality scores and modified bases were analyzed by generating pileup tables from aligned reads using samtools mpileup (v1.21)[43] with the options "--ff REVERSE,SECONDARY,QCFAIL,DUP --no-BAQ -a -d 8000 -Q 1 --no-output-ins --no-output-del", and ONT modkit pileup (v0.4.4; https://github.com/nanoporetech/modkit) with the options "--header --max-depth 1000000". Data visualization was performed using the Python libraries pandas[44], seaborn[45] and matplotlib[46]. To enable direct comparison between transcript variants, sequence breaks corresponding to construct-specific truncations or extensions were introduced, as indicated in **Extended Data Fig. 1**. To examine per-read modification quality values reported by Dorado, ONT modkit extract full was used with the options "--mapped-only --num-reads 10000".

Identification of modified bases on mutated IVT RNA reads with m6Anet and SWARM

To assess the presence of m^6A modifications on individual reads with alternative methods, we have used *de novo* modification-calling approaches, m6Anet (2.1.0)[31] and SWARM[22]. Both tools require aligned ONT signal data as one of the inputs. Input preparation was performed according to the respective tool guidelines. In both workflows, signal alignment of sorted out reads of transcript variants was carried out using f5c eventalign (v1.6)[47] with shared options: "--signal-index --scale-events –rna --kmer-model --slow5". For m6Anet, eventalign table output was used, whereas for SWARM, SAM output with read names was generated, including the recommended additional options "--min-mapq 0 --secondary=yes –samples". Aligned events were subsequently processed using m6Anet dataprep and SWARM_read_level.py with default parameters to generate per-read tables of modification probabilities. The reported modification probabilities distributions were visualized.

Comparison of signal (current) level and dwell time

To compare ionic current levels and dwell times, raw ONT signal data were aligned to the reference genome using uncalled4[48] with the options "--kit SQK-RNA004 --flowcell FLO-PRO004RA --min-aln-length 60 --eventalign-out". To plot the data we took a random sample of 250 reads. For event level, model level and event length we show the median of the data with 95% confidence interval as reported by seaborn lineplot. The comparsion of event level with model level was computed as a difference of means at each position and the Z-confidence interval of difference of means at α=0.05 was computed by CompareMeans zconfint_diff from statsmodels package[49].

Transcriptome wide analysis of low-QS regions

Low-QS regions within *B. subtilis* transcriptome were identified from non-modified IVT DRS data. Reads were pre-processed with fastplong[50] to trim low-quality ends (-5 -3 -W 5 -M 30) and aligned to the reference genome using minimap2 with "-ax map-ont" settings. Strand specific coverage and QS statistics were computed from aligned reads using samtools mpileup shared options "--no-BAQ -a -d 0 -Q 1 --no-output-ins --no-output-del". Forward strand pileups were generated using "--ff REVERSE,SECONDARY,QCFAIL,DUP", and reverse strand pileups using "--rf REVERSE --ff SECONDARY,QCFAIL,DUP". Low-QS positions were identified from the resulting pileups by computing local Z-score of mean QS values in 5-bp window relative to a 100-bp surrounding window. The positions satisfying the local Z-score <= -2 and coverage at least 100 were grouped such the positions were within 5bp of each other and the position with the lowest Z-score was reported as group representative position. Positions overlapping annotated non-coding RNAs were excluded from further analysis. To select the low QS positions, we took group representative position with mean lower quartile QS less than 12.

Analysis of secondary structures upstream of low-QS regions

A total of 1,946 positions with the lowest per-base quality scores (QS) were selected as described above. For each position, ±200 nt of flanking sequence were extracted. RNA secondary structure was analyzed using 30-nt sliding windows with a 5-nt step size. For each window, the minimum free energy (ΔG) structure was predicted using

RNAfold (v2.7.2)[51], and the number of base pairs in the predicted structure was recorded. Nucleotide frequencies (A, C, U and G) were computed in parallel. All values were averaged across positions and plotted as a function of window position relative to the low-QS site.

Low-QS regions motif analysis

To assess whether low-QS regions are associated with conserved sequence motifs, we performed de novo motif discovery using STREME [32] with settings for RNA(-rna); second-order Markov model (--order 2), and central distance objective function (-objfun cd). The minimum motif width was set to 6 nucleotides (--minw 6), and analyses were performed with three maximum motif widths (--maxw 8, 12, and 16) to capture motifs of varying lengths.

Prediction of locally stable secondary structures

Locally stable RNA secondary structures were predicted using RNALfold (ViennaRNA package v2.7.2)[51] with the options --noLP and -L 30, restricting predictions to locally optimal structures with a maximum base-pair span of 30 nucleotides. Structures with a predicted minimum free energy $\Delta G < -15$ kcal mol^{-1} were retained, and 20-nt upstream sequence regions were extracted for downstream analysis. Predictions and sequence extraction were performed separately for both strands.

Colocalization of false modifications with low-QS regions and secondary structures

To assess colocalization between low-QS regions and modified sites, we defined each low-QS region as a ±10 bp window around the detected low-quality position. Modified positions were called using modkit pileup (v0.4.4) with options "--header --edge-filter 8"; from the resulting pileups, only sites with valid coverage ≥100 reads and a modification fraction >10% were retained for downstream analyses. The same filtering setting was used to plot modification profiles in Fig. 6. Colocalization/enrichment of modified sites within low-QS windows or upstream of secondary structures were quantified using Fisher's exact test as implemented in bedtools fisher (v2.31.1)[52] with same strand enforcement (-s)[52]. Because we consider the strands separately, we take twice the genome length as the chromosome size parameter.

Further, to plot the percentage of modified bases around selected genomic regions we have used modkit localize (v0.4.4) with input data prepared by modkit pileup (see above). Plot centers were defined as follows: for low-QS regions (Fig. 6g), the centers of the low-QS regions; for predicted RNA secondary structures (Fig. 6h), the –1 position from the predicted structure 5' end, and for intrinsic transcription terminators (Fig. 6i) the –1 position from the terminator 5' end. Coordinates of intrinsic transcriptional terminators were obtained from published data[53] and mapped to *B. subtilis* BSB1 genome. For the bottom panels, control distributions were generated by randomizing the corresponding plot centers using bedtools shuffle.

Cross-organism examples of structure-linked artefacts (human and E. coli)

Publicly available DRS datasets from *E. coli* (BioProject NCBI accession PRJNA1198641)[30] and the human HEK293T cell line (ENA accession PRJEB80229)[29] were manually inspected together with their corresponding in IVT control RNAs. Reads were basecalled using Dorado model v5.2.0 with modified-base detection enabled, and aligned to the respective reference genomes with mimimap2 integrated with Dorado (human GRCh38/hg38 mm2-opts "-x splice", *E. coli* ASM75055v1 accession CP009273.1 mm2-opts "-x map-ont")[42]. Aligned reads were visualized in Integrative Genomics Viewer (IGV)[28] and colored by either sequencing errors (mismatches/indels) or predicted modifications. Genomic loci exhibiting enrichment of basecalling errors and modification calls in both IVT and *in vivo* samples (suggesting that these signals are independent of cellular modification machinery) were selected for structural analysis. Putative downstream RNA secondary structures together with their minimum free energy (MFE) were predicted by RNAfold (v2.7.2)[51] using default parameters.

Acknowledgements

This research has been supported by the Ministry of Education, Youth and Sports of the Czech Republic grant RNA for therapy (CZ.02.01.01/00/22_008/0004575) to SN and VK and by Czech Science Foundation No. 24-10700S to JW. The research was also supported by the Czech Academy of Sciences—Strategy AV21, grant number VP40. This work was carried out with the support of ELIXIR CZ Research Infrastructure (ID LM2023055, MEYS CR) and the support of the Genomics and Bioinformatics service facility of Institute of Molecular Genetics (supported by RVO, ID 68378050) of the Czech Academy of Sciences which provided access to Oxford Nanopore Technologies sequencing equipment.

Data Availability

Sequencing data are available at ENA under accession PRJEB111622. Other data and code used for the analysis are available at Zenodo under accession 10.5281/zenodo.19605487.

Author contributions

MS contributed data curation and implemented the bioinformatic processing and analysis pipelines, and performed the majority of downstream bioinformatic analyses. JP contributed transcriptome-wide RNA secondary-structure analysis. SN contributed preparation of DNA templates and *in vitro* transcription of RNA constructs. VK contributed structure analysis and figure generation. JW contributed study conceptualization, performed the DRS experiments, and contributed data analysis and interpretation. All authors contributed to discussions and to manuscript writing and editing.

Competing interests

The authors declare no competing interests.

Materials & Correspondence

Correspondence and requests for materials should be addressed to JW (email: jana.wiedermannova@biomed.cas.cz)

1. Boccaletto, P. *et al.* MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* **46**, D303–D307 (2018).
2. Gilbert, W. V. & Nachtergaele, S. mRNA regulation by RNA modifications. *Annu. Rev. Biochem.* **92**, 175–198 (2023).
3. Motorin, Y. & Helm, M. tRNA stabilization by modified nucleotides. *Biochemistry* **49**, 4934–4944 (2010).
4. Yu, J. *et al.* Dynamic m6A modification regulates local translation of mRNA in axons. *Nucleic Acids Res.* **46**, 1412–1423 (2018).
5. Delaunay, S. & Frye, M. RNA modifications regulating cell fate in cancer. *Nat. Cell Biol.* **21**, 552–559 (2019).
6. Xu, X. *et al.* The emerging clinical application of m6A RNA modification in inflammatory bowel disease and its associated colorectal cancer. *J. Inflamm. Res.* **14**, 3289–3306 (2021).

7. Sun, L. *et al.* Decoding m6A mRNA methylation by reader proteins in liver diseases. *Genes Dis.* **11**, 711–726 (2023).

8. Hu, B. *et al.* Rewired m6A of promoter antisense RNAs in Alzheimer's disease regulates neuronal genes in 3D nucleome. *Nat. Commun.* **16**, 5251 (2025).

9. Huang, Y. *et al.* M6A methylation modification in autoimmune diseases, a promising treatment strategy based on epigenetics. *Arthritis Res. Ther.* **25**, 189 (2023).

10. Hewel, C. *et al.* Direct RNA sequencing enables improved transcriptome assessment and tracking of RNA modifications for medical applications. *Nucleic Acids Res.* **53**, 53 (2025).

11. Cruciani, S. & Novoa, E. M. The new era of single-molecule RNA modification detection through nanopore base-calling models. *Nat. Rev. Mol. Cell Biol.* **27**, 10–18 (2025).

12. Meyer, D. *et al.* Unlocking the full potential of nanopore sequencing: tips, tricks, and advanced data analysis techniques. *Nucleic Acids Res.* **54**, 23 (2026).

13. Mears, M. C., Read, Q. D. & Bakre, A. Comparison of direct RNA sequencing of Orthoavulavirus javaense using two different chemistries on the MinION platform. *J. Virol. Methods* **333**, 115103 (2025).

14. Workman, R. E. *et al.* Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* **16**, 1297–1305 (2019).

15. Diensthuber, G. *et al.* Enhanced detection of RNA modifications and read mapping with high-accuracy nanopore RNA basecalling models. *Genome Res.* **34**, 1865 (2024).

16. Liu-Wei, W. *et al.* Sequencing accuracy and systematic errors of nanopore direct RNA sequencing. *BMC Genomics* **25**, 528 (2024).

17. Xie, S. *et al.* A dual context-aware basecaller for nanopore direct RNA sequencing. *Nat. Commun.* **17**, 1851 (2026).

18. Jones, J. D. *et al.* Integrating mass spectrometry with Nanopore direct RNA sequencing for de novo modification profiling of bacteriophage MS2. *bioRxiv* 2026.06.01.729303 (2026) doi:10.64898/2026.06.01.729303.

19. Wang, Y., Zhao, Y., Bollas, A., Wang, Y. & Au, K. F. Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* **39**, 1348 (2021).

20. Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **20**, 129 (2019).

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

22. Prodic, S. *et al.* SWARM: A Single-Molecule Workflow for High-Precision Profiling of RNA Modifications. *bioRxiv* 2025.12.18.695332 (2026) doi:10.64898/2025.12.18.695332.

23. Pratanwanich, P. N. *et al.* Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. *Nat. Biotechnol.* **39**, 1394–1402 (2021).

24. Leger, A. *et al.* RNA modifications detection by comparative Nanopore direct RNA sequencing. *Nat. Commun.* **12**, 7198 (2021).

25. Stephenson, W. *et al.* Direct detection of RNA modifications and structure using single-molecule nanopore sequencing. *Cell Genom.* **2**, 100097 (2022).

26. Wilson, K. S. & Von Hippel, P. H. Stability of Escherichia coli transcription complexes near an intrinsic terminator. *J. Mol. Biol.* **244**, 36–51 (1994).

27. Peters, J. M., Vangeloff, A. D. & Landick, R. Bacterial transcription terminators: the RNA 3'-end chronicles. *J. Mol. Biol.* **412**, 793–813 (2011).

28. Robinson, J. T. *et al.* Integrative Genomics Viewer. *Nat. Biotechnol.* **29**, 24 (2011).

29. Zou, Y. *et al.* A comparative evaluation of computational models for RNA modification detection using nanopore sequencing with RNA004 chemistry. *Brief. Bioinform.* **26**, bbaf404 (2025).

30. Guo, Z. *et al.* Enhanced detection of RNA modifications in Escherichia coli utilizing direct RNA sequencing. *Cell Rep. Methods* **5**, 101168 (2025).

31. Hendra, C. *et al.* Detection of m6A from direct RNA sequencing using a multiple instance learning framework. *Nat. Methods* **19**, 1590–1598 (2022).

32. Bailey, T. L. STREME: accurate and versatile sequence motif discovery. *Bioinformatics* **37**, 2834–2840 (2021).

33. Riquelme-Barrios, S. *et al.* Direct RNA sequencing of the Escherichia coli epitranscriptome uncovers alterations under heat stress. *Nucleic Acids Res.* **53**, gkaf175 (2025).

34. Zhong, Z. D. *et al.* Systematic comparison of tools used for m6A mapping from nanopore direct RNA sequencing. *Nat. Commun.* **14**, 1906 (2023).

35. Jain, M., Abu-Shumays, R., Olsen, H. E. & Akeson, M. Advances in nanopore direct RNA sequencing. *Nat. Methods* **19**, 1160 (2022).

36. Spealman, P., Burrell, J. & Gresham, D. Inverted duplicate DNA sequences increase translocation rates through sequencing nanopores resulting in reduced base calling accuracy. *Nucleic Acids Res.* **48**, 4940–4945 (2020).

37. Zhang, Z. *et al.* Systematic calibration of epitranscriptomic maps using a synthetic modification-free RNA library. *Nat. Methods* **18**, 1213–1222 (2021).

38. Maestri, S. *et al.* Benchmarking of computational methods for m6A profiling with Nanopore direct RNA sequencing. *Brief. Bioinform.* **25**, 1–13 (2024).

39. Nicolas, P. *et al.* Condition-dependent transcriptome reveals high-level regulatory architecture in Bacillus subtilis. *Science* **335**, 1103–1106 (2012).

40. Bouras, G. *et al.* Hybracter: enabling scalable, automated, complete and accurate bacterial genome assemblies. *Microb. Genom.* **10**, 001244 (2024).

41. Tatusova, T. *et al.* NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**, 6614–6624 (2016).

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

43. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

44. Mckinney, W. Data Structures for Statistical Computing in Python. *Proc. Python Sci. Conf.* **9**, 56–61 (2010).

45. Waskom, M. L. seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).

46. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

47. Gamaarachchi, H. *et al.* GPU accelerated adaptive banded event alignment for rapid comparative nanopore signal analysis. *BMC Bioinformatics* **21**, 343 (2020).

48. Kovaka, S. *et al.* Uncalled4 improves nanopore DNA and RNA modification detection via fast and accurate signal alignment. *Nat. Methods* **22**, 681–691 (2025).

49. Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. *Proc. Python Sci. Conf.* **9**, 92–96 (2010).

50. Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* **2**, e107 (2023).

51. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms for Molecular Biology* **6**, 26 (2011).

52. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

53. Mandell, Z. F. *et al.* Comprehensive transcription terminator atlas for Bacillus subtilis. *Nat. Microbiol.* **7**, 1918–1931 (2022).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- RNAfoldinginDRSSupplementarytable1.pdf
- RNAfoldinginDRSSupplementarytable2.pdf
- RNAfoldinginDRSextendeddata.docx