An epitranscriptomic map of RNA modifications on integrated HPV18 expressed from HeLa cells

Ida Rye Hellebek[1], **Johanna Jönsson**[1,2], **Jamie Yam Auxillos**[1,2], **and Sarah Rennie**[1,✉]

[1]Section for Computational and RNA Biology, Department of Biology, Copenhagen University, Ole Maaløs Vej 5, Copenhagen, Denmark
[2]Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Denmark.

Human papillomavirus(HPV)-driven cancers remain a major global health burden, and understanding how regulatory mechanisms fine-tune viral gene expression is critical for developing new therapeutic strategies. Here, we reanalyse publicly available data from the HeLa cell line to map the epitranscriptomic landscape of HPV18 transcripts. We combine Oxford Nanopore direct RNA-sequencing reads from both native RNAs and in-vitro transcribed controls to identify a robust set of m^{6}A sites supported by multiple tools, which we validate using an orthogonal chemical-profiling assay. We further show that the modification rates at many of these sites are reduced upon knockdown or inhibition of components of the m^{6}A writer-complex, indicating their dependence on the host methyltransferase machinery. We also uncover a cluster of m^{6}A marks upstream of the E6*I splice donor, which were strongly enriched in unspliced transcripts, implicating m^{6}A in regulation of E6 intron retention. In contrast, we find no convincing evidence of m^{5}C or pseudouridine in HPV18 transcripts. As the first high-resolution epitranscriptomic map of HPV18, our study highlights a potential mechanistic link between m^{6}A modification and viral RNA processing.

epitranscriptomics | HPV | cancer | Nanopore direct RNA-sequencing

Correspondence: *sarah.rennie@bio.ku.dk*

Introduction

Human papillomavirus (HPV) is a small, non-enveloped DNA virus with a circular genome of approximately 8 kb in length (Zhang et al., 2025). HPV is divided into high- and low-risk types based on their association with cancer (Doorbar et al., 2012). Of the various HPV types, HPV18 and HPV16 are the two most prevalent high-risk types, together accounting for 70% of HPV-related cervical cancers worldwide, as well as a rising fraction of other cancers such as oropharyngeal carcinomas (de Martel et al., 2017). Despite the availability of vaccines and cervical screening programmes, HPV-associated cancers remain a global health burden, underscoring the need for continued research (Sung et al., 2021; Schiffman et al., 2016).

HPV-driven oncogenesis depends on the expression of the viral E6 and E7 oncoproteins. E7 is required to drive the host cells back into cell cycle re-entry, leading to unscheduled DNA synthesis and cell proliferation, while E6 targets the tumour suppressor p53 for degradation, enabling the infected cell to evade apoptosis and become immortalised (Doorbar et al., 2012). Later in the viral infection process, accumulation of the E2 protein leads to a down-regulation of E6 and E7 expression and results in the expression of late genes (Doorbar et al., 2012). However, In the majority of long-term HPV infections, E2 is truncated or lost, resulting in the continuous expression of E6 and E7 (Nishimura et al., 2000; Molina et al., 2024). For example, modern HeLa cells harbour multiple tandem copies of integrated HPV18 with large portions of the viral genome (including E2 and late genes) disrupted, leading to viral-host chimeric transcripts that utilize a host polyadenylation signal (Yu et al., 2024). These viral transcripts often read through the E1 coding region, which encodes the viral replicate helicase, and whose expression in HeLa has been observed to affect the expression of immune response genes (Castillo et al., 2014). The E6/E7 mRNA is bicistronic and contains an intron within the E6 open reading frame; splicing of this intron generates the E6*I isoform, which is required for efficient translation of E7. This is achieved by introducing a frame-shift that in turn disrupts E6, such that the unspliced transcript is also needed in smaller quantities to maintain E6 protein expression (Tang et al., 2006).

Gene expression is heavily regulated post-transcriptionally by a diverse set of covalent chemical modifications which affect RNA nucleotides, impacting processing, stability and translation (Delaunay et al., 2024; Roundtree et al., 2017).

In messenger RNA, the most prevalent internal modification is N6-methyladenosine (m^6A), which typically occurs at the consensus motif DRACH (D=A/G/U, R = A/G, H= A/C/U). m^6A is deposited onto RNA by the "writer" core complex, including METTL3 which catalyses the transfer of the methyl group to the adenine base, and METTL14 which assists in substrate RNA recognition and positioning. This process is assisted by WTAP or other factors (e.g. VIRMA, RBM15, ZC3H13) which are associated with the core complex (Ping et al., 2014). These marks are additionally read by reader proteins, that have a large influence on RNA fate (Zaccara et al., 2019), the most well-known being the YTH-domain family, with others, such as members of the IGF2BP family, acting as additional readers (Huang et al., 2018). In addition to m^6A, two further modifications occurring on mRNA are 5-methylcytosine (m^5C) and pseudouridine (Ψ). These are installed by their own writer enzymes (e.g. NSUN or DNMT2 for m^5C, PUS family enzymes for Ψ) and have been widely implicated in regulating RNA processing and translation (Song et al., 2022; Liu et al., 2022; Borchardt et al., 2020).

Viruses are able to exploit the molecular machinery of the host cells to regulate viral transcription via RNA modification. For example, m^6A plays a critical role in the replication cycle of human immunodeficiency virus (HIV-1) (Lichinchi et al., 2016). In the context of HPV, it was recently shown that the HPV E7 oncogene mRNA is subject to m^6A methylation and that this mark increases E7 mRNA stability by recruiting the m^6A-binding protein IGF2BP1 (Wang et al., 2022). These findings highlight the importance of RNA modifications in shaping virus–host interactions and raise the question of how epitranscriptomic marks influence the regulation of HPV transcripts and oncogenic potential.

Whilst transcriptome-wide detection of precise modified positions has historically been challenging, the recent development of third-generation sequencing has enabled the direct mapping of RNA modifications by sequencing full-length native RNA molecules. Nanopore-based direct RNA sequencing (DRS) detects changes in ionic current as RNA passes through the pore (Leger et al., 2021; Garalde et al., 2018). As the ionic current signal is altered at in the presence of chemical base modifications, specialised bioinformatic tools can be used to map modifications on a per-molecule basis (Pratanwanich et al., 2021; Kovaka et al., 2025; Gamaarachchi et al., 2020; Hendra et al., 2022). In practice, samples incorporating *in-vitro* transcribed (IVT) mRNA (which lacks modifications) or knocked down writers are often employed as controls for avoiding false positives (Tan et al., 2024; Acera Mateos et al., 2024; Zou et al., 2025), and many studies additionally rely on a combination of approaches to reliability identify sites (Zhong et al., 2023). Recent high-throughput chemical based assays have also been developed to profile specific modifications. For example, GLORI allows for detection and quantification of m^6A sites by chemically converting unmodified adenosines to inosine (Liu et al., 2023), and for m^5C, m^5C-TAC-seq was recently developed (Lu et al., 2024).

In this study we aimed to map the precise modified locations of post-transcriptional RNA modifications in HeLa-derived HPV18 transcripts. Using a number of public-available datasets, including a recent DRS dataset in HeLa, we identified a robust set of m^6A sites supported by multiple tools and assays. Taking advantage of the long-read nature of direct RNA sequencing, we identify a cluster of m^6A sites in the viral early region, upstream of the E6 intron donor site, that exhibit significantly higher modification levels in unspliced transcripts than in spliced transcripts, suggesting a potential role for m^6A in modulating splicing at this locus. Finally, we find that there is little evidence for m^5C or pseudouridine in HPV18 transcripts based our analyses under the conditions tested, although future improvements in detection sensitivity may reveal additional modifications.

Results

Direct RNA-sequencing data provides sufficient read coverage of the HPV18 genome

To assess RNA modifications on HPV18 transcripts, we re-analysed a recently published direct RNA sequencing (DRS) dataset generated from HeLa cells (Zou et al., 2025). This dataset was selected for two key reasons. First, it was generated using Oxford Nanopore's newer RNA004 flow cell, which provides improved throughput and signal quality compared to earlier versions. Second, it includes both a native (wild-type, WT) and an *in-vitro* transcribed (IVT) control, where the IVT sample serves as an unmodified baseline. Our analysis followed the workflow outlined in Figure 1A, using the Oxford Nanopore Dorado basecaller with modification-aware models to detect m^6A, m^5C, and pseudouridine (Ψ), and aligning the reads to a hybrid reference genome (human GRCh38 plus HPV18).

The HPV18-mapped read lengths formed two clusters (approximately 0.8 kb and 2.3 kb) for both IVT and WT samples (Figure 1B). The shorter reads most likely correspond to E6/E7 transcripts while the longer reads to transcripts incorporating E6, E7, and E1 regions. Almost all reads contained soft-clipped human sequences visible in IGV (Figure S1A), which we excluded from downstream analyses. By examining the presence of viral exons and 5' end start positions of each read, we classified the reads into six splice variant (SV) categories, which were present in both IVT and native read sets (Figure 1C). Five correspond to isoforms identified in a previous study (Yu et al., 2024), while a sixth (SV6) was newly defined here, retaining both the E6 intron and the E1 region. The median read length for each splice variant was consistent with expectations (Figure 1D). Furthermore, the coverage profile across

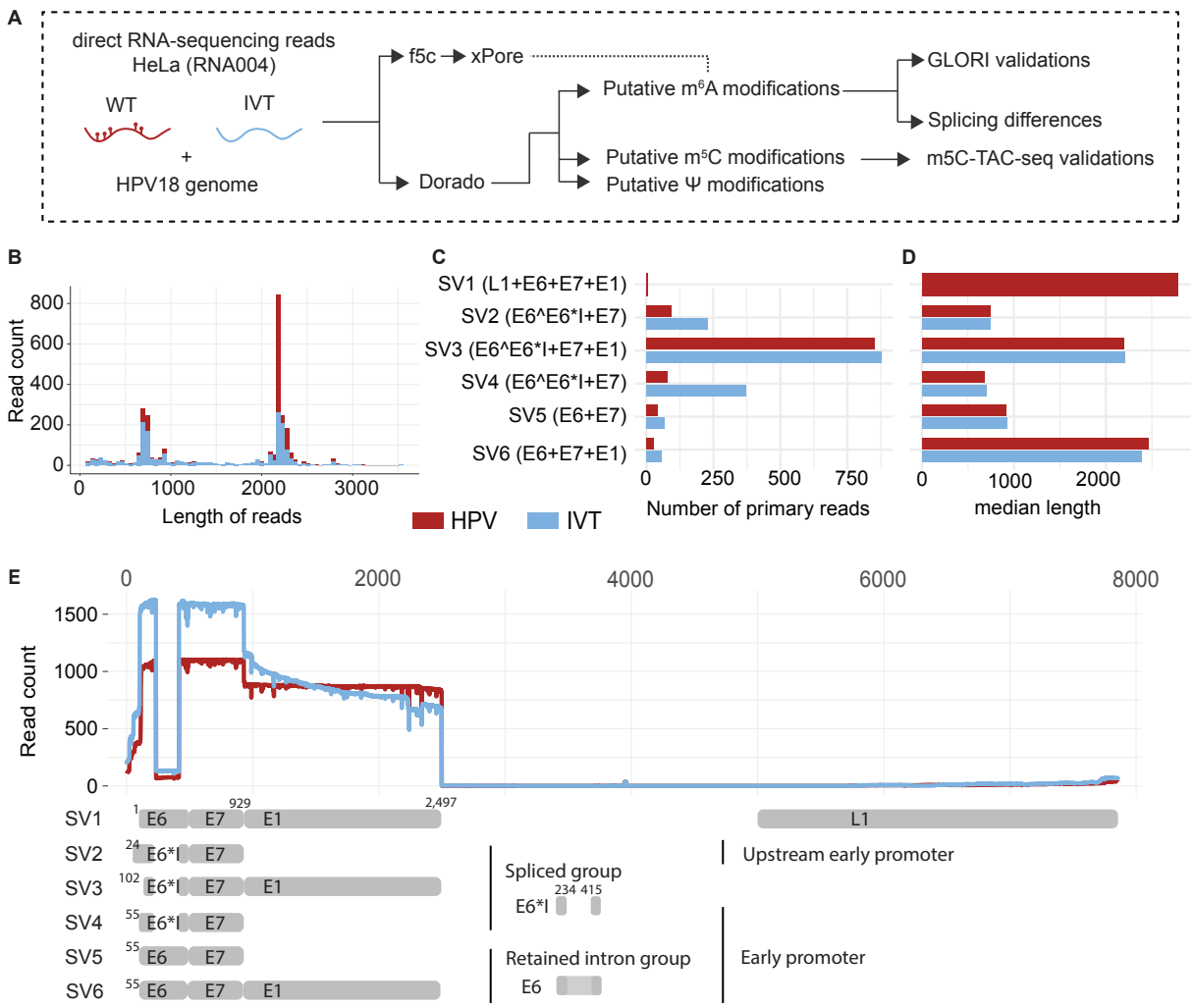

Figure 1. Nanopore direct RNA-sequencing (DRS) reads map to the HPV18 genome. (**A**) Schematic of the overall study design and analysis workflow. (**B**) Length distribution of DRS reads mapping to the HPV18 genome, shown separately for the WT and IVT samples. (**C**) Number of reads mapping to each of six splice variant (SV) categories. (**D**) Median read lengths for each splice variant category. (**E**) Coverage across the HPV18 genome by nanopore reads, including schematic illustrating the six splice variant categories.

the HPV18 genome confirmed high read depth over the E6 and E7 coding regions but low coverage elsewhere (Figure 1D), and the quality of the data was generally good in these high-coverage regions (Figure S1B).These analyses indicate that DRS reads map in sufficient quantity and quality to the HPV18 genome, supporting their use for downstream detection and characterisation of RNA modification sites on viral transcripts.

Dorado identifies multiple candidate m^{6}A sites on HPV18 transcripts

We next extracted the modification status per position aggregated across reads in the HPV18-mapped reads. Dorado reported numerous candidate sites for each of the three queried modification types (m^{6}A, m^{5}C, and Ψ) in both the WT and IVT samples (Figure 2A). Out of the three, only the m^{6}A calls had substantially more candidate sites in the WT sample than in the IVT sample. By contrast, the numbers of candidate m^{5}C or Ψ sites in IVT were similar to (or higher than) those found in native RNA from WT, implying that most of those calls are likely false positives rather than genuine modifications.

Focussing on m^{6}A, there were 25 detected candidate sites identified in the WT HPV18 transcripts, whereas only 6 sites were called in the IVT sample (Figure 2A), of which 4 positions were found in both (Figure S2A). In support of the WT candidates, we observed that the number of sites detected per individual molecule was much greater for the WT than the IVT controls (Figure S2B), and that they typically occurred in common variations of the DRACH motif (Figure 2B). Next, we compared the m^{6}A stoichiometries for each site between WT and IVT samples, where stoichiometry is defined here as the proportion of modified reads (based on probability cutoffs from the tool Modkit by Oxford Nanopore Technologies) (Figure 2C). This analysis revealed a subset of sites with substantially higher modification

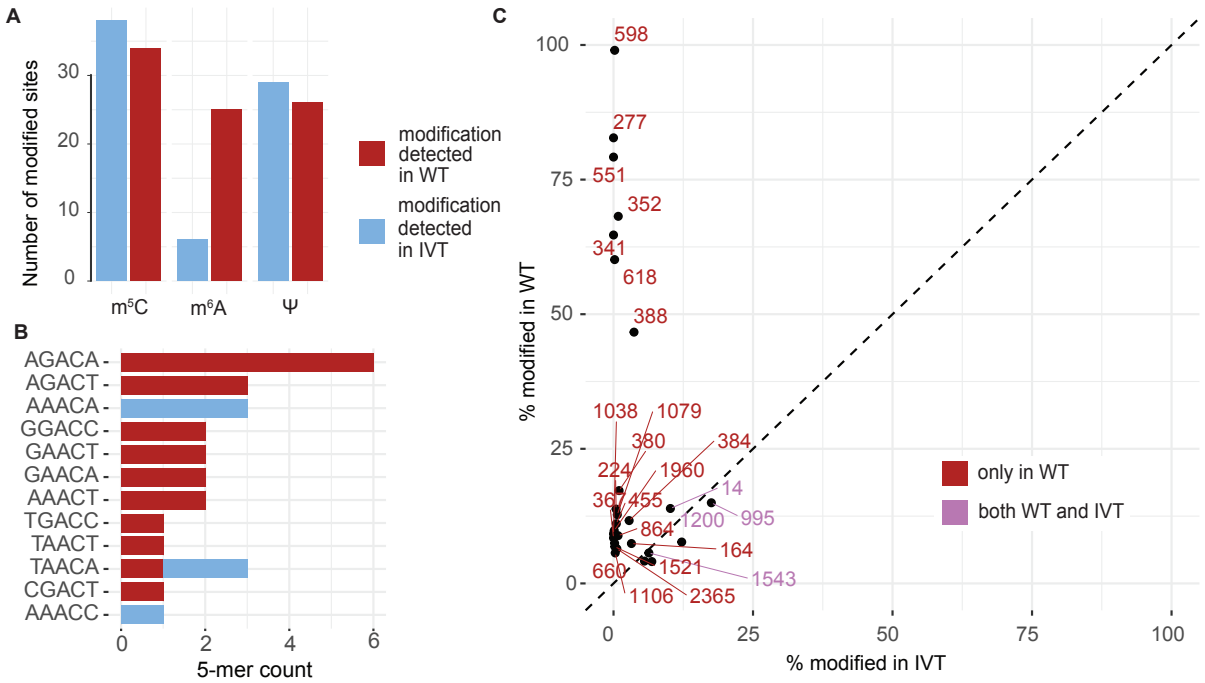

Figure 2. m^{6}A modification supported by Dorado. (**A**) Number of high-confidence sites called by Dorado in HPV18, for each of m^{5}C, m^{6}A and pseudouridine, and split by WT or IVT. (**B**) 5-mers centered on the putative m^{6}A sites (**C**) Scatter plot showing modification rates between WT and IVT for Dorado identified sites. Red depicts sites with a modification rate greater than 5% in WT, purple depicts sites with a modification rate greater than 5% in both samples.

stoichiometry in WT native state, suggesting that these sites may be true m^{6}A sites in the viral transcripts.

Cross-validation of m^{6}A sites using xPore

Given that Dorado's models may not have been trained on viral sequences, we sought independent support for the candidate m^{6}A sites using an alternative, data-driven approach. We applied the tool xPore (Pratanwanich et al., 2021), which uses Gaussian mixture modelling to detect statistically significant differences in raw ionic current signals at each nucleotide between two conditions. Upon applying xPore to compare the IVT sample and the WT sample, we experienced ambiguities in labelling which of the two inferred states was modified or unmodified using the software. However, given the IVT sample represents an unmodified ground state, we focussed on the absolute difference in modification rate, ignoring the sign (Figure 3A). The analysis found 25 candidate positions, from which only three sites overlap with the 25 sites identified by Dorado (Figure 2A,C). However, it has previously been described how signal leakage and differences in the most influential base within the pore can cause the signal aligned to neighbouring nucleotides to appear more greatly perturbed by a modification (Kovaka et al., 2025; Makhamreh et al., 2024). When we considered the 5-basepair window (+/-2 bp) of each candidate, we found an additional 12 positions which agreed between Dorado and xPore (Figure 3A), which was significantly greater than what would be expected by random chance (Figure S2C). Notably, none of the significant xPore positions occurred near the four Dorado sites that were also detected in the IVT control, meaning that 15 out of the 21 (71%) non-IVT detected positions in WT were supported by xPore.

In addition, we examined event-alignment plots from f5c (Gamaarachchi et al., 2020) for two of the positions, 277 and 551 (Figure 3B,C). For position 277, the WT and IVT signals appeared to overlap at the nucleotide itself, but there was a clear separation one base upstream (position 276) (Figure 3B), aligning with xPore calling the guanine at position 276 (Figure 3A). Given that both 276 from xPore and 277 from Dorado had similar modification rates (around three-quarters), we conclude it very likely supports the m^{6}A site at 277. In contrast, at position 551 the WT and IVT signals appear distinctly separated at the modified base itself (Figure 3C). We conclude that, although xPore's site assignments can be shifted due to signal dispersion or other technical factors, its analysis supports many of the m^{6}A sites identified by Dorado.

Orthogonal evidence for m^{6}A from GLORI data

We next sought to validate the 25 candidate m^{6}A sites identified by DRS using the orthogonal chemical profiling method, GLORI, which identifies m^{6}A sites based on their selective protection from A-to-I conversion (Liu et al.,

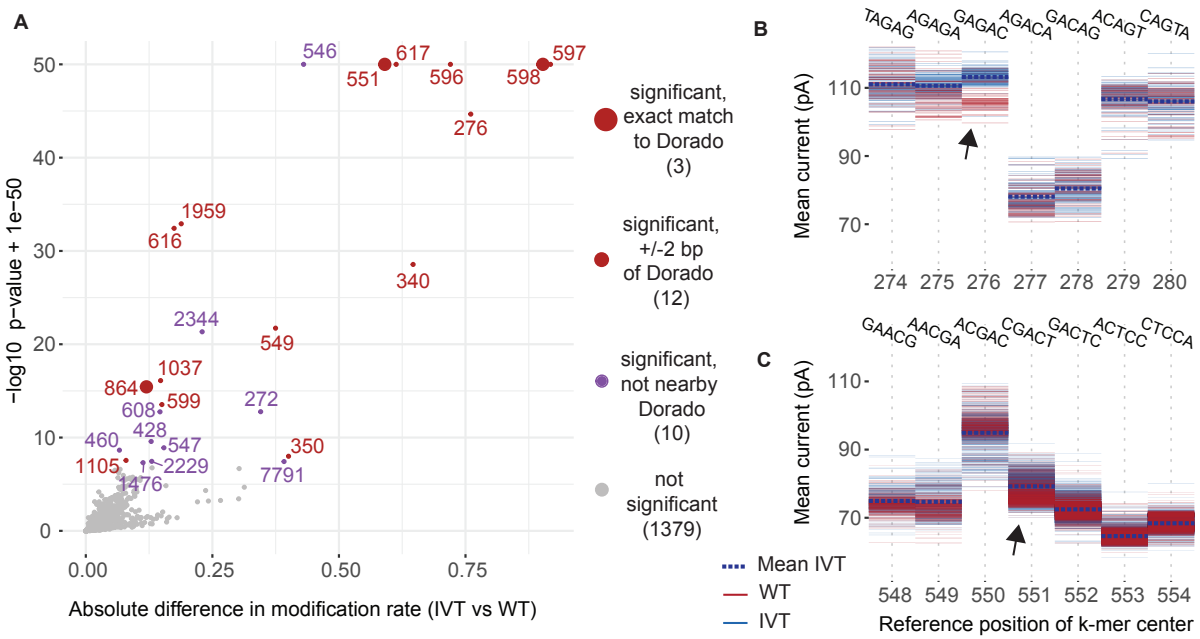

Figure 3. xPore analysis supports majority of Dorado-identified m^{6}A sites. (**A**) xPore results comparing WT and IVT samples. Each point represents a genomic position, plotted by xPore's inferred modification rate difference (WT–IVT, absolute value) versus -log10(adjusted p-value). Points are coloured by xPore significance (purple for FDR < 0.05, grey for not significant). Red circles denote positions corresponding to Dorado m^{6}A candidates (large red circles for exact nucleotide matches; small red circles for positions within ±2 nt of a Dorado-called site). (**B-C**) f5c eventalign signal plots for WT (red) and IVT (blue) reads around two Dorado-identified m^{6}A candidate sites, with the position identified by xPore is indicated by black arrows: (**B**) position 277 and (**C**) position 551, including 3bp up/downstream from the site of interest. Thick dotted horizontal lines are the mean signal level for the IVT samples, providing a reference reflecting the unmodified state.

2023). We reprocessed all relevant GLORI sequencing datasets from HeLa cells (including various control and perturbation conditions, see Methods) with the HPV18 genome added to the reference. We first examined the pooled GLORI control group (combined data from multiple chemically-treated control samples for each of the perturbation experiments), which provided substantial read coverage across the HPV18 transcriptome (Figure S3A). For each of the 25 Dorado-called candidate sites (Figure 2A), we calculated the fraction of adenosines that remained unedited ("protected" As) (Figure 4A). Many sites showed protected fractions consistent with the modification stoichiometries estimated from direct RNA sequencing data, while IVT-supported sites showed protection levels only marginally different from than the observed background based on the analogous samples not treated by the GLORI (A-to-I conversion) procedure (Figure S3B). Overall, for the 21 WT-specific m^{6}A candidates, the GLORI-derived methylation fractions correlated strongly with the nanopore-based estimates (r^2=0.96, Figure S3C). Furthermore, after stringent filtering procedures (see Methods), we identified 9 additional sites in GLORI with a total editing proportion of at least 5% across the GLORI-treated control samples (Figure 4B). It is likely that these sites were not detected in Dorado for two reasons. First, some may have been filtered out from the Dorado-detection workflow, either because they have inherently low editing proportions, or because they do not sit in the DRACH sequence context expected by Dorado's m^{6}A-detection model (Figure 4B). For example, position 932 has a modification rate of 21% according to GLORI and sits in a GTACA motif, where the T in second position is not typically found in the canonical DRACH motif. Second, many of them occur in regions where the coverage by Nanopore DRS was extremely low (Figure 4C), meaning that it would not be possible to find them, if they are true modification sites.

Next, we reanalysed the GLORI perturbation experiments to assess the dependence of these m^{6}A sites on the host methylation machinery. These experiments involved siRNA knockdowns of the m^{6}A writer complex components (METTL3, METTL14 and WTAP), knockdown of a reader protein (YTHDF2), or the chemical treatment with a METTL3 inhibitor (STM2457). Comparing each perturbation condition with its respective control revealed that many of the m^{6}A sites showed decreased methylation upon disruption of the writer complex, particularly under METTL3 inhibition or WTAP knockdown (Figure 4D, Figure S4A). Here, we note that not all sites could be tested due to low read coverage in the relevant samples. For example, position 598 is consistently almost 100% methylated according to results from both direct RNA sequencing and GLORI, but dropped to around ∼34% upon METTL3 inhibition, and ∼73% upon WTAP knockdown, with a similar pattern observed for position 551. Interestingly, position 455, despite being low-stoichiometry, was also significantly reduced upon these perturbations (Figure 4D) and two of the additional positions 932 and 687 also had significantly reduced modification rate upon METTL3 inhibition and WTAP

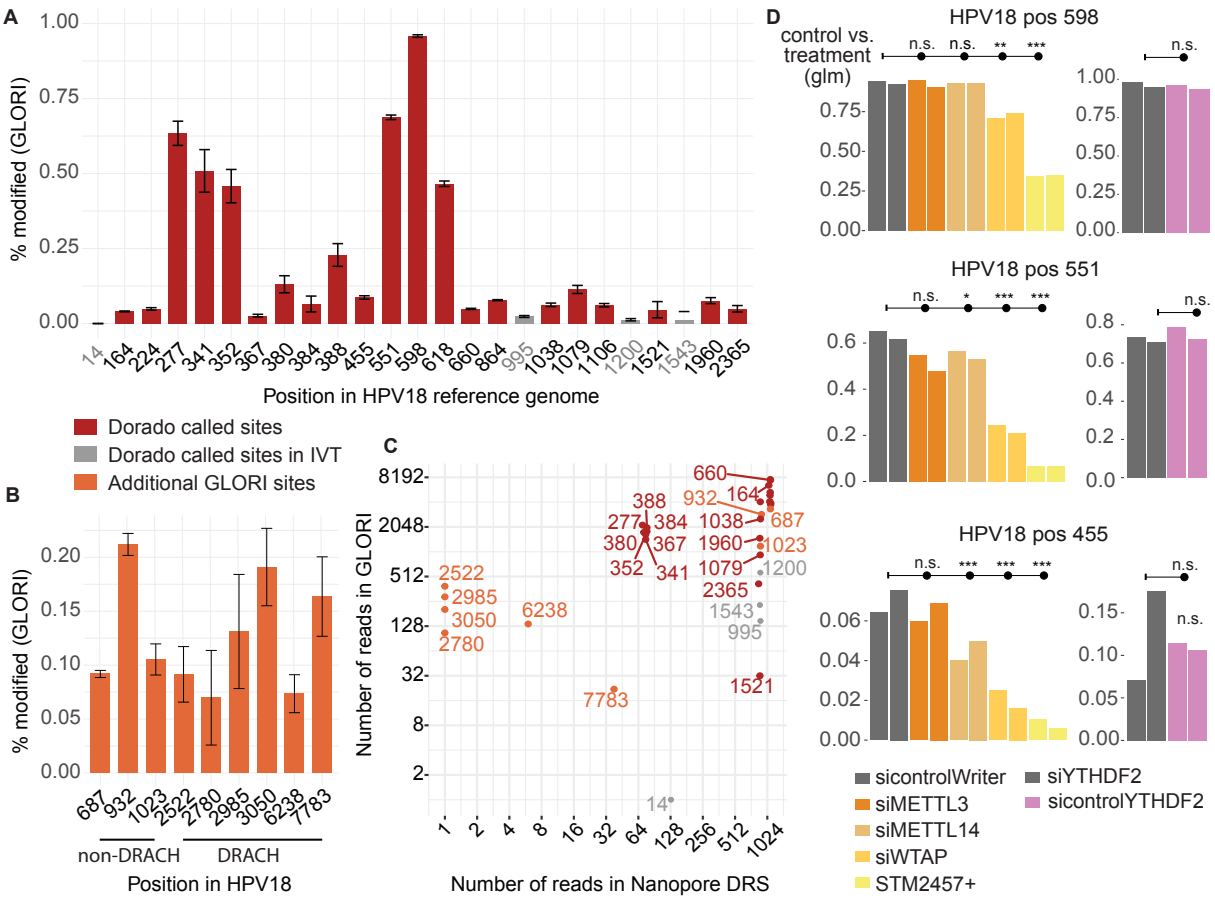

Figure 4. Concordance between DRS and GLORI for HPV18 m^{6}A sites. (**A**) Fraction of adenosines protected from A-to-I conversion (m^{6}A fraction) at each of the 25 Dorado-identified m^{6}A candidate positions, as measured by GLORI in pooled control samples. Grey marked positions represent the 4 candidates which were also detected in the IVT sample. (**B**) Identification of nine additional m^{6}A sites by GLORI. (**C**) Number of reads covering all candidate positions from (A) and (B) for DRS and GLORI, coloured according to whether they are identified in DRS WT-only (red), both WT and IVT (grey) or GLORI-only (orange). (**D**) Candidate sites 598, 551 and 455 in controls and in the various perturbations of the m^{6}A writer complex, as well as before and after siRNA knockdown of YTHDF2. Dual bars per position represents the two replicates from GLORI. Bars above represent the significance of each perturbation, based on binomial logistic regression models, compared to the grey control samples. Key: n.s. Not significant, * p-value<0.05, ** p-value<0.01 *** p-value<0.001.

knockdown (Figure S4A). In contrast, none of the 25 candidates from direct RNA-sequencing or the additional GLORI candidates showed a significant change in methylation upon YTHDF2 knockdown relative to controls (Figure 4D, Figure S4A), suggesting that loss of this reader protein does not appreciably affect the presence of the methylation marks. Additionally, some sites show altered m^{6}A levels under hypoxic conditions (Figure S4B). Taken together, we conclude that HPV18 m^{6}A sites are effectively captured by GLORI and provide additional support to sites called by the DRS data, and that their installment appears, at least in-part, to be dependent on the host methyltransferase writer complex.

Evidence for splicing-associated m^{6}A sites upstream of E6*I intron donor

One advantage of long-read direct RNA sequencing is the ability to assess RNA modifications in the context of specific splice variants. Given the well-defined splicing event of HPV18 leading to the E6*I transcript, we asked whether m^{6}A stoichiometries differed between unspliced and spliced variants. Since E6 intron splicing is highly efficient, we pooled our six defined splice variants into two groups, to maximise statistical power: those containing full length E6 (the unspliced group, 62 reads) versus E6*I transcripts (the spliced group, 1005 reads) (Figure 1E, Figure 5A). Focussing on the early region ∼1-900 bp of the HPV18 genome, we observed several patterns (Figure 5B). First, consistent with earlier observations, Dorado clearly identified a set of m^{6}A sites which were specific to the intron (Figure 5B, center), with the strongest modification rate occurring at candidate closest to the 5' splice donor (position 277).

Most intriguingly, we observed five adenosine sites in the exonic region upstream of the intron (positions 58, 164, 168,

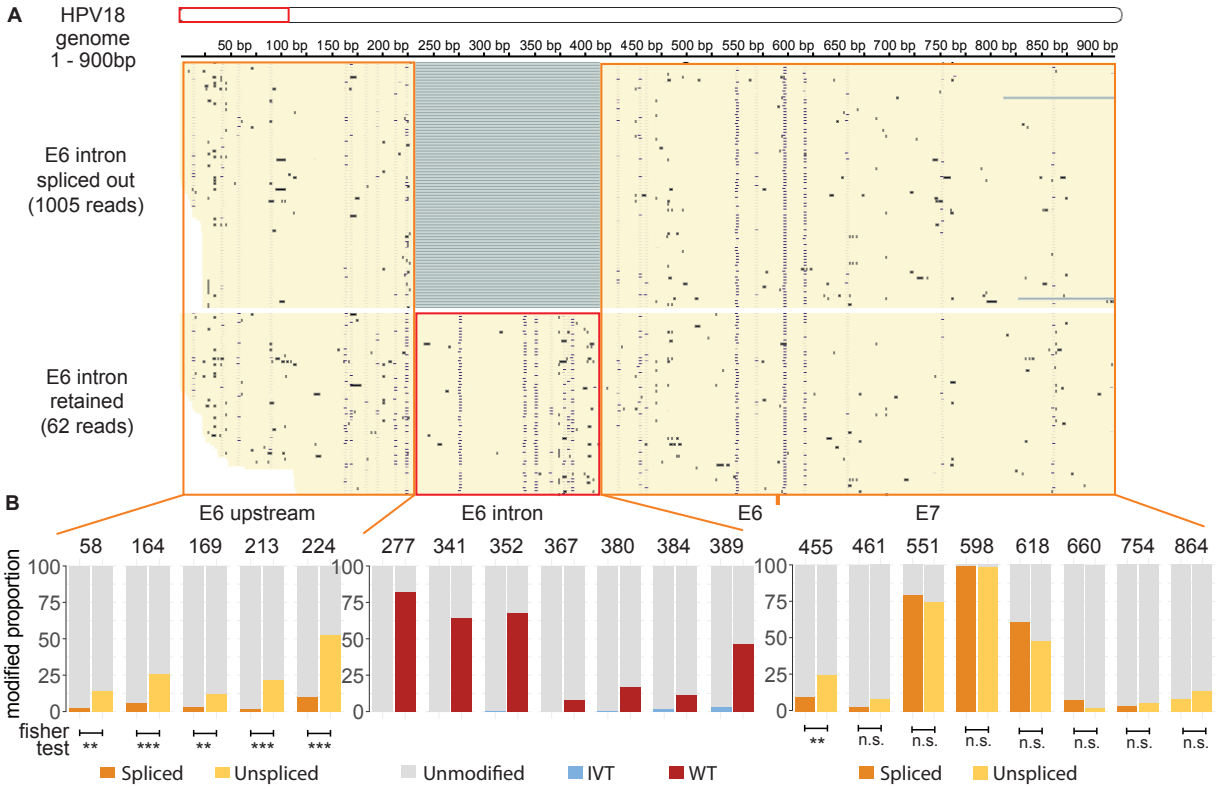

Figure 5. E6 intron splicing status influences m^{6}A levels on HPV18 transcripts. (**A**) IGV browser view of approx. ∼1-900 nt within the HPV18 genome, split into two tracks according to reads either containing (unspliced) or not containing (spliced) the E6 intron. Blue dots represent m^{6}A modification scores as determined by Dorado. (**B**) Bar plots for 20 candidate m^{6}A sites with significantly higher methylation in WT versus IVT (Fisher exact test). Fisher test for left and right groups represents testing for differences in m^{6}A stoichiometry in spliced versus unspliced transcripts. Key: n.s. Not significant, * p-value<0.05, ** p-value<0.01 *** p-value<0.001.

213 and 224) which showed significantly higher m^{6}A stoichiometry in unspliced reads compared to spliced reads (Figure 5B, left, Fisher Exact test on counts of spliced/unspliced and modified/unmodified). This difference was the most prominent at position 224 (p=3.96 e-10), an identified m^{6}A site sitting in a GAACT sequence context just upstream of the splice donor site. Due to the predominance of spliced reads (∼93%), this position only had ∼14% stoichiometry in previous analyses, however in unspliced reads, the methylation rate reached ∼53% methylation rate. In comparison to the upstream region, we annotated 8 candidate m^{6}A sites downstream of the E6 intron acceptor site. Apart from the site at position 455, which is a AAACT context 40bp downstream of the E6 intron acceptor site (p=5.05e-03), none of these downstream positions showed statistical evidence of differential stoichiometry in the unspliced vs. spliced read sets. Importantly, none of the 20 putative m^{6}A sites considered in this analysis showed appreciable modification in the IVT control (even when stratified by isoform), indicating the differences are not due to sequencing artifacts (Figure S5A).

To further support the link between intron retention and methylation at site 224, we turned to the GLORI data. Although GLORI's short-read approach cannot directly assign modifications to full-length isoforms, we devised a proxy: we extracted all reads mapping to position 224 and then created two subsets – those spanning the exon-exon junction (indicating the spliced E6*I) or those covering the intron region (full-length E6). While noisy, this analysis supported the trend of significantly higher inferred m^{6}A at position 224 in unspliced reads (Figure S5B). The overall low protected A ratio in this analysis is likely due to ambiguities in allocating reads to the unspliced forms rather than a disagreement with the direct RNA-sequencing analyses.

As an additional analysis, we also asked whether perturbation of the m^{6}A host machinery affected the relative portions of spliced vs unspliced reads in GLORI (by counting the number of reads overlapping at least in part the E6 intron, per sample). We observed a slight increase in intron retention upon METTL3 knockdown (Figure S5C), consistent with a recent report that METTL3/METTL14 depletion increases HPV18 intron retention (Cui et al., 2022). Finally, we explored whether other splicing or transcript features might be associated with differential m^{6}A deposition. In particular, we examined (1) transcripts that include the E1 region versus those that do not, and (2) transcripts initiating from an earlier cryptic initiation site at +24 versus others (SV2 versus others, Figure 1E, Yu et al. (2024)), to see if any m^{6}A sites showed isoform-specific patterns. In both cases, we did not detect clear differences in m^{6}A

levels analogous to the E6 intron-retention scenario (Figure S6A, S6B). In summary, our results reveal a distinct cluster of m^{6}A marks that are preferentially present on transcripts retaining the E6 intron, pointing to an association between m^{6}A modifications and E6 intron retention in HPV18.

Integrated analysis highlights 8 high-confidence m^{6}A sites linked to the canonical writer complex

Finally, we integrated evidence from both the nanopore and GLORI analyses in order to pinpoint the most confident m^{6}A sites on HPV18 transcripts. To this end, we prioritised sites that were 1) detected by DRS only in the WT sample (and not in IVT) and 2) showed a significant decrease in methylation upon at least one m^{6}A-writer perturbation in the GLORI data. Table 1 lists the 8 m^{6}A sites meeting these strict criteria, whose methylation levels range from 7% to 99% in the direct RNA-seq dataset, although this does not preclude many of the other sites identified in our analyses. Notably, three of these high-confidence sites (positions 224, 455 and 164) also showed splicing-dependent methylation patterns (based on our above analyses), further reinforcing the link between RNA splicing and dynamic m^{6}A deposition on HPV18 transcripts.

Table 1. Top 8 m^{6}A sites supported by direct RNA-sequencing (WT only) and significantly perturbed by at least one condition from the GLORI dataset. Positions are annotated according to: 1-based coordinate in the HPV18 genome, DRS% is the stoichiometry determined from Nanopore, GLORI% is the proportion of protected A from GLORI, Significant conditions is the GLORI conditions significantly perturbed (p<0.01, binomial GLM), E6 intron splicing associated means that the position has splicing-dependent m^{6}A levels (p<0.01, Fisher Exact test).

Position	DRS %	GLORI %	Significant conditions (GLORI)	E6 intron splicing associated?
598	99	96	STM2457$^+$,siWTAP	no
551	79	69	STM2457$^+$,siWTAP	no
618	60	47	STM2457$^+$,siWTAP,siMETTL3	no
224	14	5	siWTAP	yes
455	11	9	STM2457$^+$,siWTAP	yes
864	9	8	STM2457$^+$,siWTAP,siMETTL14	no
660	7	5	STM2457$^+$,siWTAP	no
164	7	4*	STM2457$^+$,siWTAP	yes

* Under the 5% cutoff used in this study to identify m^{6}A sites from GLORI data.

Little evidence for m^{5}C or pseudouridine modifications in HeLa-integrated HPV18 transcripts

As shown earlier, we observed that Dorado detected a number of 5-methylcytosine (m^{5}C) sites in both WT and IVT samples on HPV18 RNAs (Figure 2A). Of the 34 sites identified in the WT sample, 26 of them were also found in the IVT sample (Figure 6A), raising suspicion that most were false positives. When comparing the stoichiometry between WT and IVT samples for the sites detected in WT, we noticed that the sites tended to group around the identity line, with only a few with the potential to have higher stoichiometry in the WT samples (Figure 6B, highlighted in red). Interestingly, we noticed that three of these WT-only m^{5}C calls were located exactly one nucleotide downstream of three of the strongest m^{6}A sites (that is, a C following the DRACH A of site 551, 598 and 618, Figure 6C). These raised the possibility that the presence of a nearby m^{6}A might have confounded the m^{5}C model, due to the altered signal current caused by m^{6}A at the adjacent upstream base. This is supported by the fact that the three calls were not predicted in the IVT data (where the m^{6}A is also absent) (Figure 6C, lower track).

To directly test whether HPV18 transcripts carry m^{5}C, we re-analysed an orthogonal sequencing dataset, m^{5}C-TAC-seq, for HeLa cells (Lu et al., 2024). In m^{5}C-TAC-seq, methylated cytosines are ultimately detected as C-to-U mutations resulting from the TET-oxidation and selection strategy. We reprocessed the HeLa WT and IVT control samples and looked for cytosine positions on HPV18 with evidence of significant conversion ratios relative to those observed in either input and IVT control samples. Special attention was given to the eight m^{5}C sites detected by Dorado as specific to the WT direct RNA sequencing sample, all which had reasonable coverage in the m^{5}C-TAC-seq samples (Figure S7A). This analysis showed only weak conversion ratios in the treated WT samples, that were in all cases statistically not different from the controls (Figure 6D). In other words, the chemical labelling assay did not validate any m^{5}C on HPV18, supporting that the Dorado m^{5}C calls were indeed false positives.

Finally, in the case of Ψ, of the 26 sites detected by Dorado with at least 25 reads coverage and 5% stoichiometry, 23 of them were also detected by the IVT sample (Figure S7B). Comparing the stoichiometries between WT and IVT, all points were close to the identity line, leaving no clear candidates for further analysis (Figure S7C), and suggesting overall that there is low evidence based on our analyses of Ψ in the HPV18 transcriptome. Although we cannot rule out the presence of these modifications in regions where coverage of the direct RNA sequencing reads was very low, we conclude, based on our analyses, that there is little evidence of m^{5}C or Ψ sites in HPV18 transcripts.

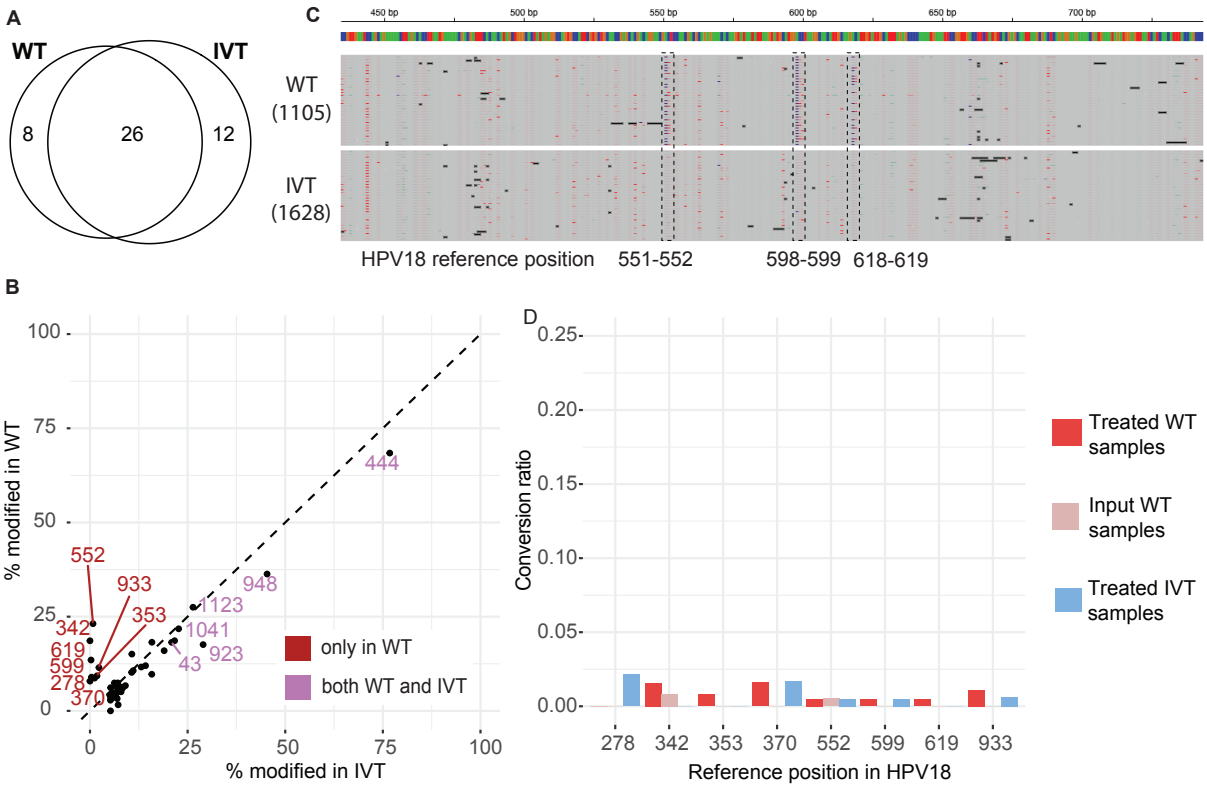

Figure 6. Minimal evidence for m^{5}C modifications on HPV18 transcripts. (**A**) Venn diagram showing overlap of Dorado-identified m^{5}C sites between the WT and IVT samples (**B**) IGV snapshot illustrating an example of a false-positive m^{5}C call. In the WT reads, Dorado calls an m^{5}C at cytosines (indicated by red marks) immediately following three of our identified m^{6}A sites (blue marks). (**C**) Scatter plot comparing m^{5}C modification stoichiometry in WT vs. IVT at all sites called in WT. (**D**) m^{5}C-TAC-seq results for the eight Dorado WT-specific m^{5}C candidate sites. No positions showed statistically significant conversion ratios compared to control samples (based on a binomial generalised linear model).

Discussion

Nanopore long-read sequencing is increasingly being applied to study HPV in cancer biology. For example, it has been used to precisely map viral integration sites and characterise viral transcript structures in HPV-positive cell lines (Parida et al., 2024). Here, we utilised direct RNA nanopore sequencing, in combination with orthogonal assays, to identify RNA modifications on integrated HPV18 transcripts in HeLa cells, creating, to the best of our knowledge, the first high-resolution epitranscriptomic map for HPV18 RNAs. We identified a set of high-confidence m^{6}A sites on HPV18 transcripts that were consistently present in the native viral RNA whilst largely absent in unmodified IVT controls. These candidate sites were cross-validated by two computational strategies and additionally showed concordant modification levels in an independent chemical profiling assay. In contrast, our analyses found very little evidence for two other internal modifications (notably m^{5}C or pseudouridine) on the HPV18 mRNAs, although we cannot exclude the presence of low-stoichiometry sites or sites in regions poorly covered by this analysis. We discuss more specific findings in the context of relevant literature below.

First, viruses are known to "hijack" RNA host modification machinery, allowing them to incorporate m^{6}A on their mRNAs, with recent literature reporting strong effects on viral replication and gene expression. In HPV, a recent report by Wang et al. (2022) demonstrated that m^{6}A on the E7 mRNA increases its stability via the reader IGF2BP1. Consistent with their results, we identified a major m^{6}A site in the E7 region of HPV18 (position 598) that is almost fully methylated under normal conditions. The methylation at this site is highly sensitive to METTL3 inhibition and WTAP knockdown, supporting it as a bona fide m^{6}A site written by the canonical METTL3/METTL14/WTAP methyltransferase complex. Given the high sequencing homology between HPV16 and HPV18 E7, it is likely that this site corresponds to the one implicated in the IGF2BP1-mediated stabilisation of E7. If so, we speculate that position 598 may be critical for maintaining the high levels of E7 oncogene required for its proliferation and survival, and suggest that further work to elucidate the role of this m^{6}A site in HPV18 oncogene stabilisation would be warranted.

Our study also illustrates the utility of combining DRS with orthogonal validation methods for mapping RNA modifications. DRS provides single-molecule resolution of modification profiles and links modifications to specific transcript isoforms. However, DRS alone is prone to false positives and can show variable sensitivity depending on sequence context. By using two different computational detection tools (Dorado and xPore), which rely on distinct algorithms, and then cross-checking their outputs with independent methods (GLORI for m^{6}A and m^{5}C-TAC-seq for m^{5}C), we achieved a high degree of confidence in the modification sites we called. In addition, our study underscores the importance of using IVT (unmodified RNA) controls as negative references (Zou et al., 2025). The inclusion of IVT controls proved vital for distinguishing false-positive m^{6}A, m^{5}C or Ψ calls from true modifications, a practice increasingly recommended in epitranscriptomic studies in virsuses or bacteria (Tan et al., 2024; Tan et al.).

One of the most interesting findings from our analysis is an association between m^{6}A modifications and the alternative splicing of the HPV18 E6 intron. Using the long DRS reads, we observed that several adenosines upstream of the E6 intron 5' splice donor site are heavily methylated in transcripts that retain the intron, but only weakly methylated in E6*I transcripts where the intron is spliced out. Notably, the adenosine at position 224 (which lies in a GAACT motif right at the 5' splice site) showed 50% methylation in the unspliced transcripts, yet only 13% methylation when averaged across all reads (most of which were spliced). In other words, the minority-group of transcripts in which the E6 intron was not removed accumulated a much higher m^{6}A signal at this site, a finding which illustrates how bulk measurements of m^{6}A can mask important differences in rare sub-populations.

Our observations suggest a model where m^{6}A deposition and splicing of the E6 intron are interconnected. Relevant to this, recent literature reports that perturbing the m^{6}A pathway alters E6/E7 splicing in HPV16: in particular, depletion of the m^{6}A writers METTL3/METTL14 or overexpression of the m^{6}A-related factors ALKBH5 (a demethylase) or YTHDC1 (a nuclear reader) leads to increased retention of the E6 intron in HPV16 (Cui et al., 2022). These findings are consistent with more general roles of m^{6}A affecting co-transcriptional splicing (Wei et al., 2021), and our re-analysis of published GLORI data supports a similar trend in HPV18 in the case of perturbations in the m^{6}A writer components. Based on this, we might speculate a model whereby m^{6}A marks installed on pre-mRNAs help recruit YTHDC1 and other splicing factors to the E6 intron, facilitating its removal (Xiao et al., 2016; Ke et al., 2017). Once splicing is successfully completed, the assembly of the exon–exon junction complex (EJC) at that site may hinder any further m^{6}A addition in the vicinity, while the absence of the EJC in unspliced transcripts might allow them to accumulate higher m^{6}A levels due to prolonged writer access (Yang et al., 2022). Whilst clarifying these mechanisms would require detailed follow-up experiments, which is beyond the scope of the present study, our results point to an interplay between m^{6}A modification and alternative splicing of HPV18 RNAs.

In contrast to the E6 intron, we did not observe notable m^{6}A differences associated with other viral RNA processing events. For example, the inclusion of the E1 exon or choice of a cryptic transcription initiation site previously observed in (Yu et al., 2024) (+24, SV2) did not show differential m^{6}A patterns in our data. It may be that splicing-associated m^{6}A in the context of the E6 intron is the most biologically relevant effect of RNA modifications in this system. However, we note that our coverage of regions outside E6/E7 was limited, reducing our power to detect modifications

there. Indeed, our re-analysis of GLORI data hinted at a few additional m^{6}A sites at low stoichiometry on HPV18 transcripts that were not detected in the DRS data. These could be interesting targets for future analysis, especially on late transcripts or in other sample conditions. Additionally, we note that nanopore reads often truncate at the 5' end, which could obscure our ability to confidently distinguish transcripts originating from different start sites and to compare their modification levels.

The study has additional limitations to acknowledge. Our analysis was based on a single replicate for each condition (WT and IVT), due to the constraints of available public data. Whilst the newer RNA004 chemistry provides improved yields, read coverage was still limited for some transcripts, particularly the less abundant unspliced transcripts, as discussed above. Replicating this analysis in additional direct RNA sequencing Nanopore datasets would be recommendable. Second, the biological context we examined is a cervical cancer cell line (HeLa) harbouring an integrated HPV18 genome that is known to be truncated and rearranged, meaning that our findings may not fully generalise to infections with intact HPV18. Despite these limitations, our study provides a useful starting point for understanding RNA modifications in HPV-driven cancers and generates testable hypotheses for future research in patient samples or additional models.

Conclusion

Using two orthogonal datasets and multiple analysis methodologies, we generated a high-confidence map of m^{6}A modifications in HPV18 transcripts using direct RNA sequencing. Our findings suggest that m^{6}A modifications may confer an advantage to the HPV18 virus by facilitating the robust expression of E6/E7 oncoviral mRNAs. As a first map of epitranscriptomic marks on HPV18 RNA, our work lays a foundation for future functional studies on how RNA modifications might regulate HPV gene expression in the context of cancer.

Methods

All downstream analyses of processed data were conducted in R using custom scripts, unless otherwise stated. All plots, unless otherwise stated, were generated using the R package ggplot2. IGV browser was used to inspect reads and generate screenshots (Thorvaldsdóttir, Robinson, and Mesirov 2013). P-values were adjusted using Benjamini-Hochberg procedure (False Discovery Rate), unless otherwise stated.

Direct RNA sequencing analysis with Dorado

Dorado raw data processing: Direct RNA-sequencing data for HeLa and associated IVT was downloaded from (Zou et al., 2025). Dorado tool kit (v0.9.0) was used to basecall raw pod5-files with the model `rna004_130bps_sup@v5.1.0`. While basecalling, reads were aligned to a reference consisting of HPV18-transcripts (from Yu et al. (2024)) merged with transcripts from GRCh38 (Gencode v47), using minimap2 with default settings as provided by Dorado. Following, reads with a primary alignment to a HPV transcript with mapping quality > 5, were remapped to the HPV18-genome using minimap2 (v2.28) with options `-ax splice -uf -k14 -t 16 -y --secondary=no`.

Dorado modification calling: Basecalling was also done with the available modification models `m6A_DRACH`, `pseU` and `m5C`. Predicted modification probabilities were subsequently analysed with Modkit (v0.4.2) by Oxford Nanopore, using `modkit pileup` with default options to filter, tabulate and calculate position-level stoichiometries, whereas `modkit extract full` were used to calculate modified sites per molecule. Only positions detected from Dorado with at least 5% stoichiometry and at least 25 reads mapping, were retained for further analyses.

Read summaries: Read lengths, median lengths and coverage were computed excluding soft-clipped regions, using samtools, `dorado summary` and custom bash scripts.

Splice variants: Reads primarily aligned to the HPV18 were assigned 6 splice variants (SV1-6) using bedtools according to 1) 5' end coordinate of read 2) splicing of E6 intron (180-184N in CIGAR string) 3) presence of E1 sequence and read ids were used to create 6 separate .bam files for each of the WT and IVT samples. For the analysis of spliced versus unspliced reads, SV2-5 were merged (spliced) and SV 5-6 were merged (unspliced). For the analysis of presence versus absence of E1 SV,3 and 6 were merged (present) and SV 2,4,5 were merged (absent). All bam files were inspected using IGV browser before continuing with downstream analyses.

Analysis of spliced vs. unspliced read sets: for each site, a Fisher Exact test was performed on a 2 x 2 table of counts (either spliced/unspliced or WT/IVT compared with modified/unmodified). Only positions significantly more modified WT vs IVT (adjusted p-value <0.01) were retained for assessing the significance of spliced vs. unspliced.

Sequence analysis: The R package GenomicRanges combined with BSgenome and the reference for the HPV18 genome in order to extract the 5-mer context at the candidate m^{6}A sites.

Direct RNA sequencing analysis with xPore

Eventalign: All eventalign files were generated with the Nanopolish re-implementation f5c (v1.5), to allow for analysis of RNA004 data. The raw data was converted from pod5 to blow5-format using blue-crab (v0.3.0), and then indexed with f5c index. Reads mapped to HPV18 (basecalled with Dorado as previously described) were converted to fastq using samtools. Reference genome, blow5-files, fastq-files and corresponding bam-files were all passed to f5c eventalign together with the options `-signal-index`, `--scale-events`, `--rna`, `--min-mapq 0`, `--kmer-model rna004.nulceotide.5mer.model`.

xPore: Using the f5c-generated eventalign file, xPore (v2.0) was run with default settings except the RNA004 k-mer table that was downloaded from xPore github repository (https://raw.githubusercontent.com/GoekeLab/xpore/RNA004_kmer_model/xpore/diffmod/RNA004_5mer_model.txt) and supplied as "prior" in the config.yml before running xpore diffmod.

Downstream filtering: As xPore does not handle genetic variation, all positions with +/- 4 nt of a SNP, were removed from the xPore output. SNPs with > 10% frequency were extracted from basecalled reads using samtools pileup. Additionally, the positions inferred as modified were also filtered based on p-values (p < 1e-7) suggested in another publication (Tan et al.). P-value adjustment in Figure 3A is designed to break 0-defined p-values solely for the purposes of plotting.

GLORI data analyses

Data preprocessing: raw sequencing reads for all GLORI samples in HeLa (both GLORI+, for samples involving the chemical protocol, and GLORI- samples not treated by the chemical protocol) were downloaded from NCBI using SRA-tools (Leinonen et al., 2011). For the GLORI+ samples, a version of the reference genome for GRch38 + HPV18 was created with computationally converted As to Gs and indexed using STAR (STAR-2.7.11a). Reads were mapped using STAR, with options `--clip3pAdapterSeq (AGATCGGAAGAGC)`, `--clip3pAfterAdapterNbases 0`, `--outFilterMismatchNoverReadLmax 0.10`, `--outFilterMismatchNmax 999`, `--outFilterMultimapNmax 50`, `--outFilterScoreMinOverLread 0`, `--outFilterMatchNminOverLread 0`, `--chimSegmentMin 15`, `--chimOutType WithinBAM SoftClip`, `--sjdbOverhang 74` and `--outSAMattributes NH HI AS nM NM MD`. Mapped reads were filtered for a minimum mapping quality score of 20, filtered for alignment to HPV18 and samtools mpileup (Li et al., 2009) was used on all positions in the HPV18 reference genome.

Identification of HPV18 sites: For all adenosines in the HPV18 reference, counts of G (converted) and A (protected) for GLORI+ samples were used to calculate the proportion of protected As (A_count/(A_count+G_count)). For GLORI (not treated by the chemical procedure) samples the background A-to-G conversion ratio (G_count/(A_count+G_count)) was calculated. For assessing modified proportions at candidate sites in either GLORI+ or GLORI- samples, counts were first pooled across replicates of control (non-perturbed, WT) samples for the writer knockdown or inhibition experiments, the YTHDF knockdown examples or the hypoxia treatments (labelled siwritercontrol, siYTHDFcontrol and hypoxia-, respectively). Individual positions were significant if they had at least 5% protected A ratio across pooled control samples, at least 5% in 2 replicates of at least one of the same experiment and had negligible A-to-G mutation rates in the GLORI- samples. The variance at each position was calculated using np(1-p) where n is the pooled read coverage at the given position, p is the proportion of protected As.

Analysis of knockdown, inhibition or hypoxia experiments: Control samples for the m^6A writer complex perturbation experiments (siwritercontrol) were compared to each perturbation (siMETTL3, siMETTL14, siWTAP or STM2457+), the YTHDF knockdown samples (siYTHDF) were paired with their control samples (siYTHDFcontrol) and hypoxia treated samples (hypoxia+) were paired with the hypoxia controls (hypoxia-). To assess positions with significantly reduced protected-A ratios upon perturbation, a generalised linear model specifying a binomial family and condition (control or perturbed) as a covariate was used to calculate the significance of sites, using information from both replicates for each condition. Adjusted p-values were converted to -log10(.) scale prior to producing the heatmap using R package pheatmap.

Supplemental splicing analyses: Reads overlapping position 224 were extracted from the mapped bam files and were further split according to the presence of 180-184N in the CIGAR string (indicating the spliced E6 intron) or a match to the exon-exon junction (indicating the unspliced form). Samtools mpileup was used on the resulting split bam files and protected A ratios calculated as above.

Intron retention analysis: GLORI- non-treated samples were used as a basis to extract reads covering the E6 intron via inspection of alignment blocks in CIGAR strings in bam files containing HPV18-mapped reads.

m^{5}C-TAC-seq analysis

Data processing: Raw sequenced reads were downloaded using SRA-toolkit. Low quality ends and adapter sequences were trimmed from the reads using cutadapt (Martin 2011). The R2 reads were then deduplicated according to UMI (Smith et al., 2017) and clipped 10-nt, using seqkit (Shen et al., 2016) and fastp (Chen et al., 2018) respectively. The inserts of R2 were mapped to the combined GRCh38 and HPV18 genomes using hisat-3n (Zhang et al., 2021). Mismatches for C to T were filtered from mapped bam files and used to create a mutation table using hisat-3n-table command. Downstream analyses: A custom R script was used to process and analyse the resulting C to T conversion ratios. A binomial generalised linear model using the model formula `cbind(converted,unconverted) ~ RNA_type * treatment` where RNA_type is WT or IVT reads and treatment is either input or chemically treated samples, and the interaction term is used to assess a significant conversion ratio in WT chemically treated samples.

Data Availability

Nanopore direct RNA sequencing data for HeLa was downloaded from ENA database accession PRJEB80229, raw GLORI sequencing data was downloaded from NCBI Gene Expression Omnibus (GEO) (GSE210563), raw m^{5}C-TAC-seq data was downloaded from GEO (GSE242724). The reference sequence for the HPV18 genome was downloaded from NCBI (Reference Sequence: NC_001357.1).

Author contributions

I.R.H analysed the data with support from S.R, S.R. supervised the project with input from J.J and J.Y.A. S.R. wrote the manuscript with input from all authors.

Funding

S.R. is supported by the Novo Nordisk Foundation (NNF24SA0092867), J.Y.A acknowledges support by Villum Fonden (VIL58880).

Bibliography

Acera Mateos, P., J Sethi, A., Ravindran, A., Srivastava, A., Woodward, K., Mahmud, S., Kanchi, M., Guarnacci, M., Xu, J., W S Yuen, Z., Zhou, Y., Sneddon, A., Hamilton, W., Gao, J., M Starrs, L., Hayashi, R., Wickramasinghe, V., Zarnack, K., Preiss, T., Burgio, G., Dehorter, N., E Shirokikh, N., and Eyras, E. (2024). Prediction of m6A and m5C at single-molecule resolution reveals a transcriptome-wide co-occurrence of RNA modifications. *Nature Communications*, 15(1):3899. doi: 10.1038/s41467-024-47953-7. Publisher: Nature Publishing Group.

Borchardt, E. K., Martinez, N. M., and Gilbert, W. V. (2020). Regulation and Function of RNA Pseudouridylation in Human Cells. *Annual Review of Genetics*, 54:309–336. doi: 10.1146/annurev-genet-112618-043830.

Castillo, A., Wang, L., Koriyama, C., Eizuru, Y., Jordan, K., and Akiba, S. (2014). A systems biology analysis of the changes in gene expression via silencing of HPV-18 E1 expression in HeLa cells. *Open Biology*, 4(10):130119. doi: 10.1098/rsob.130119.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics (Oxford, England)*, 34(17):i884–i890. doi: 10.1093/bioinformatics/bty560.

Cui, X., Nilsson, K., Kajitani, N., and Schwartz, S. (2022). Overexpression of m6A-factors METTL3, ALKBH5, and YTHDC1 alters HPV16 mRNA splicing. *Virus Genes*, 58(2):98–112. doi: 10.1007/s11262-022-01889-6.

de Martel, C., Plummer, M., Vignat, J., and Franceschi, S. (2017). Worldwide burden of cancer attributable to HPV by site, country and HPV type. *International Journal of Cancer*, 141(4):664–670. doi: 10.1002/ijc.30716.

Delaunay, S., Helm, M., and Frye, M. (2024). RNA modifications in physiology and disease: towards clinical applications. *Nature Reviews Genetics*, 25(2):104–122. doi: 10.1038/s41576-023-00645-2. Publisher: Nature Publishing Group.

Doorbar, J., Quint, W., Banks, L., Bravo, I. G., Stoler, M., Broker, T. R., and Stanley, M. A. (2012). The Biology and Life-Cycle of Human Papillomaviruses. *Vaccine*, 30:F55–F70. doi: 10.1016/j.vaccine.2012.06.083.

Gamaarachchi, H., Lam, C. W., Jayatilaka, G., Samarakoon, H., Simpson, J. T., Smith, M. A., and Parameswaran, S. (2020). GPU accelerated adaptive banded event alignment for rapid comparative nanopore signal analysis. *BMC Bioinformatics*, 21 (1):343. doi: 10.1186/s12859-020-03697-x.

Garalde, D. R., Snell, E. A., Jachimowicz, D., Sipos, B., Lloyd, J. H., Bruce, M., Pantic, N., Admassu, T., James, P., Warland, A., Jordan, M., Ciccone, J., Serra, S., Keenan, J., Martin, S., McNeill, L., Wallace, E. J., Jayasinghe, L., Wright, C., Blasco, J., Young, S., Brocklebank, D., Juul, S., Clarke, J., Heron, A. J., and Turner, D. J. (2018). Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods*, 15(3):201–206. doi: 10.1038/nmeth.4577. Publisher: Nature Publishing Group.

Hendra, C., Pratanwanich, P. N., Wan, Y. K., Goh, W. S. S., Thiery, A., and Göke, J. (2022). Detection of m6A from direct RNA sequencing using a multiple instance learning framework. *Nature Methods*, 19(12):1590–1598. doi: 10.1038/s41592-022-01666-1. Publisher: Nature Publishing Group.

Huang, H., Weng, H., Sun, W., Qin, X., Shi, H., Wu, H., Zhao, B. S., Mesquita, A., Liu, C., Yuan, C. L., Hu, Y.-C., Hüttelmaier, S., Skibbe, J. R., Su, R., Deng, X., Dong, L., Sun, M., Li, C., Nachtergaele, S., Wang, Y., Hu, C., Ferchen, K., Greis, K. D., Jiang,

X., Wei, M., Qu, L., Guan, J.-L., He, C., Yang, J., and Chen, J. (2018). Recognition of RNA N6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nature Cell Biology*, 20(3):285–295. doi: 10.1038/s41556-018-0045-z. Publisher: Nature Publishing Group.

Ke, S., Pandya-Jones, A., Saito, Y., Fak, J. J., Vågbø, C. B., Geula, S., Hanna, J. H., Black, D. L., Darnell, J. E., and Darnell, R. B. (2017). m6A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. *Genes & Development*, 31(10):990–1006. doi: 10.1101/gad.301036.117.

Kovaka, S., Hook, P. W., Jenike, K. M., Shivakumar, V., Morina, L. B., Razaghi, R., Timp, W., and Schatz, M. C. (2025). Uncalled4 improves nanopore DNA and RNA modification detection via fast and accurate signal alignment. *Nature Methods*, 22(4): 681–691. doi: 10.1038/s41592-025-02631-4.

Leger, A., Amaral, P. P., Pandolfini, L., Capitanchik, C., Capraro, F., Miano, V., Migliori, V., Toolan-Kerr, P., Sideri, T., Enright, A. J., Tzelepis, K., van Werven, F. J., Luscombe, N. M., Barbieri, I., Ule, J., Fitzgerald, T., Birney, E., Leonardi, T., and Kouzarides, T. (2021). RNA modifications detection by comparative Nanopore direct RNA sequencing. *Nature Communications*, 12(1):7198. doi: 10.1038/s41467-021-27393-3. Publisher: Nature Publishing Group.

Leinonen, R., Sugawara, H., Shumway, M., and on behalf of the International Nucleotide Sequence Database Collaboration. (2011). The Sequence Read Archive. *Nucleic Acids Research*, 39(suppl_1):D19–D21. doi: 10.1093/nar/gkq1019.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16):2078–2079. doi: 10.1093/bioinformatics/btp352.

Lichinchi, G., Gao, S., Saletore, Y., Gonzalez, G. M., Bansal, V., Wang, Y., Mason, C. E., and Rana, T. M. (2016). Dynamics of the human and viral m6A RNA methylomes during HIV-1 infection of T cells. *Nature Microbiology*, 1(4):16011. doi: 10.1038/nmicrobiol.2016.11. Publisher: Nature Publishing Group.

Liu, C., Sun, H., Yi, Y., Shen, W., Li, K., Xiao, Y., Li, F., Li, Y., Hou, Y., Lu, B., Liu, W., Meng, H., Peng, J., Yi, C., and Wang, J. (2023). Absolute quantification of single-base m6A methylation in the mammalian transcriptome using GLORI. *Nature Biotechnology*, 41(3):355–366. doi: 10.1038/s41587-022-01487-9. Publisher: Nature Publishing Group.

Liu, J., Huang, T., Chen, W., Ding, C., Zhao, T., Zhao, X., Cai, B., Zhang, Y., Li, S., Zhang, L., Xue, M., He, X., Ge, W., Zhou, C., Xu, Y., and Zhang, R. (2022). Developmental mRNA m5C landscape and regulatory innovations of massive m5C modification of maternal mRNAs in animals. *Nature Communications*, 13(1):2484. doi: 10.1038/s41467-022-30210-0. Publisher: Nature Publishing Group.

Lu, L., Zhang, X., Zhou, Y., Shi, Z., Xie, X., Zhang, X., Gao, L., Fu, A., Liu, C., He, B., Xiong, X., Yin, Y., Wang, Q., Yi, C., and Li, X. (2024). Base-resolution m5C profiling across the mammalian transcriptome by bisulfite-free enzyme-assisted chemical labeling approach. *Molecular Cell*, 84(15):2984–3000.e8. doi: 10.1016/j.molcel.2024.06.021.

Makhamreh, A., Tavakoli, S., Fallahi, A., Kang, X., Gamper, H., Nabizadehmashhadtoroghi, M., Jain, M., Hou, Y.-M., Rouhanifard, S. H., and Wanunu, M. (2024). Nanopore signal deviations from pseudouridine modifications in RNA are sequence-specific: quantification requires dedicated synthetic controls. *Scientific Reports*, 14(1):22457. doi: 10.1038/s41598-024-72994-9. Publisher: Nature Publishing Group.

Molina, M. A., Steenbergen, R. D. M., Pumpe, A., Kenyon, A. N., and Melchers, W. J. G. (2024). HPV integration and cervical cancer: a failed evolutionary viral trait. *Trends in Molecular Medicine*, 30(9):890–902. doi: 10.1016/j.molmed.2024.05.009. Publisher: Elsevier.

Nishimura, A., Ono, T., Ishimoto, A., Dowhanick, J. J., Frizzell, M. A., Howley, P. M., and Sakai, H. (2000). Mechanisms of Human Papillomavirus E2-Mediated Repression of Viral Oncogene Expression and Cervical Cancer Cell Growth Inhibition. *Journal of Virology*, 74(8):3752–3760. doi: 10.1128/jvi.74.8.3752-3760.2000.

Parida, P., Mukherjee, N., Singh, A., Lewis, S., Sharan, K., Mallya, S., Singh, A., Rao, M., Higginson, D. S., Sabarinathan, R., and Damerla, R. R. Novel nanopore sequencing method for determining Human Papillomavirus integrations in tumors without the need for whole genome sequencing, (2024). URL https://www.biorxiv.org/content/10.1101/2024.10.17.618842v1. Pages: 2024.10.17.618842 Section: New Results.

Ping, X.-L., Sun, B.-F., Wang, L., Xiao, W., Yang, X., Wang, W.-J., Adhikari, S., Shi, Y., Lv, Y., Chen, Y.-S., Zhao, X., Li, A., Yang, Y., Dahal, U., Lou, X.-M., Liu, X., Huang, J., Yuan, W.-P., Zhu, X.-F., Cheng, T., Zhao, Y.-L., Wang, X., Danielsen, J. M. R., Liu, F., and Yang, Y.-G. (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Research*, 24(2):177–189. doi: 10.1038/cr.2014.3. Publisher: Nature Publishing Group.

Pratanwanich, P. N., Yao, F., Chen, Y., Koh, C. W. Q., Wan, Y. K., Hendra, C., Poon, P., Goh, Y. T., Yap, P. M. L., Chooi, J. Y., Chng, W. J., Ng, S. B., Thiery, A., Goh, W. S. S., and Göke, J. (2021). Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. *Nature Biotechnology*, 39(11):1394–1402. doi: 10.1038/s41587-021-00949-w. Publisher: Nature Publishing Group.

Roundtree, I. A., Evans, M. E., Pan, T., and He, C. (2017). Dynamic RNA Modifications in Gene Expression Regulation. *Cell*, 169(7):1187–1200. doi: 10.1016/j.cell.2017.05.045.

Schiffman, M., Doorbar, J., Wentzensen, N., de Sanjosé, S., Fakhry, C., Monk, B. J., Stanley, M. A., and Franceschi, S. (2016). Carcinogenic human papillomavirus infection. *Nature Reviews Disease Primers*, 2(1):16086. doi: 10.1038/nrdp.2016.86. Publisher: Nature Publishing Group.

Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE*, 11(10):e0163962. doi: 10.1371/journal.pone.0163962. Publisher: Public Library of Science.

Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Research*, 27(3):491–499. doi: 10.1101/gr.209601.116.

Song, H., Zhang, J., Liu, B., Xu, J., Cai, B., Yang, H., Straube, J., Yu, X., and Ma, T. (2022). Biological roles of RNA m5C modification and its implications in Cancer immunotherapy. *Biomarker Research*, 10(1):15. doi: 10.1186/s40364-022-00362-8.

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a cancer journal for clinicians*, 71(3):209–249. doi: 10.3322/caac.21660.

Tan, L., Guo, Z., Wang, X., Kim, D. Y., and Li, R. Utilization of nanopore direct RNA sequencing to analyze viral RNA modifications. *mSystems*, 9(2):e01163–23. doi: 10.1128/msystems.01163-23.

Tan, L., Guo, Z., Shao, Y., Ye, L., Wang, M., Deng, X., Chen, S., and Li, R. (2024). Analysis of bacterial transcriptome and

epitranscriptome using nanopore direct RNA sequencing. *Nucleic Acids Research*, 52(15):8746–8762. doi: 10.1093/nar/gkae601.

Tang, S., Tao, M., McCoy, J. P., and Zheng, Z.-M. (2006). The E7 Oncoprotein Is Translated from Spliced E6*I Transcripts in High-Risk Human Papillomavirus Type 16- or Type 18-Positive Cervical Cancer Cell Lines via Translation Reinitiation. *Journal of Virology*, 80(9):4249–4263. doi: 10.1128/jvi.80.9.4249-4263.2006. Publisher: American Society for Microbiology.

Wang, L., Zhan, G., Maimaitiyiming, Y., Su, Y., Lin, S., Liu, J., Su, K., Lin, J., Shen, S., He, W., Wang, F., Chen, J., Sun, S., Xue, Y., Gu, J., Chen, X., Zhang, J., Zhang, L., Wang, Q., Chang, K.-J., Chiou, S.-H., Björklund, M., Naranmandura, H., Cheng, X., and Hsu, C.-H. (2022). m6A modification confers thermal vulnerability to HPV E7 oncotranscripts via reverse regulation of its reader protein IGF2BP1 upon heat stress. *Cell Reports*, 41(4). doi: 10.1016/j.celrep.2022.111546. Publisher: Elsevier.

Wei, G., Almeida, M., Pintacuda, G., Coker, H., Bowness, J. S., Ule, J., and Brockdorff, N. (2021). Acute depletion of METTL3 implicates N^6-methyladenosine in alternative intron/exon inclusion in the nascent transcriptome. *Genome Research*, 31(8): 1395–1408. doi: 10.1101/gr.271635.120. Publisher: Cold Spring Harbor Laboratory.

Xiao, W., Adhikari, S., Dahal, U., Chen, Y.-S., Hao, Y.-J., Sun, B.-F., Sun, H.-Y., Li, A., Ping, X.-L., Lai, W.-Y., Wang, X., Ma, H.-L., Huang, C.-M., Yang, Y., Huang, N., Jiang, G.-B., Wang, H.-L., Zhou, Q., Wang, X.-J., Zhao, Y.-L., and Yang, Y.-G. (2016). Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. *Molecular Cell*, 61(4):507–519. doi: 10.1016/j.molcel.2016.01.012.

Yang, X., Triboulet, R., Liu, Q., Sendinc, E., and Gregory, R. I. (2022). Exon junction complex shapes the m6A epitranscriptome. *Nature Communications*, 13(1):7904. doi: 10.1038/s41467-022-35643-1. Publisher: Nature Publishing Group.

Yu, L., Majerciak, V., Lobanov, A., Mirza, S., Band, V., Liu, H., Cam, M., Hughes, S. H., Lowy, D. R., and Zheng, Z.-M. (2024). HPV oncogenes expressed from only one of multiple integrated HPV DNA copies drive clonal cell expansion in cervical cancer. *mBio*, 15(5):e00729–24. doi: 10.1128/mbio.00729-24.

Zaccara, S., Ries, R. J., and Jaffrey, S. R. (2019). Reading, writing and erasing mRNA methylation. *Nature Reviews Molecular Cell Biology*, 20(10):608–624. doi: 10.1038/s41580-019-0168-5. Publisher: Nature Publishing Group.

Zhang, Y., Park, C., Bennett, C., Thornton, M., and Kim, D. (2021). Rapid and accurate alignment of nucleotide conversion sequencing reads with HISAT-3N. *Genome Research*, 31(7):1290–1295. doi: 10.1101/gr.275193.120.

Zhang, Y., Qiu, K., Ren, J., Zhao, Y., and Cheng, P. (2025). Roles of human papillomavirus in cancers: oncogenic mechanisms and clinical use. *Signal Transduction and Targeted Therapy*, 10(1):44. doi: 10.1038/s41392-024-02083-w. Publisher: Nature Publishing Group.

Zhong, Z.-D., Xie, Y.-Y., Chen, H.-X., Lan, Y.-L., Liu, X.-H., Ji, J.-Y., Wu, F., Jin, L., Chen, J., Mak, D. W., Zhang, Z., and Luo, G.-Z. (2023). Systematic comparison of tools used for m6A mapping from nanopore direct RNA sequencing. *Nature Communications*, 14(1):1906. doi: 10.1038/s41467-023-37596-5. Publisher: Nature Publishing Group.

Zou, Y., Ahsan, M. U., Chan, J., Meng, W., Gao, S.-J., Huang, Y., and Wang, K. (2025). A Comparative Evaluation of Computational Models for RNA modification detection using Nanopore sequencing with RNA004 Chemistry. *bioRxiv*, page 2025.02.03.636352. doi: 10.1101/2025.02.03.636352.

Supplementary Figures

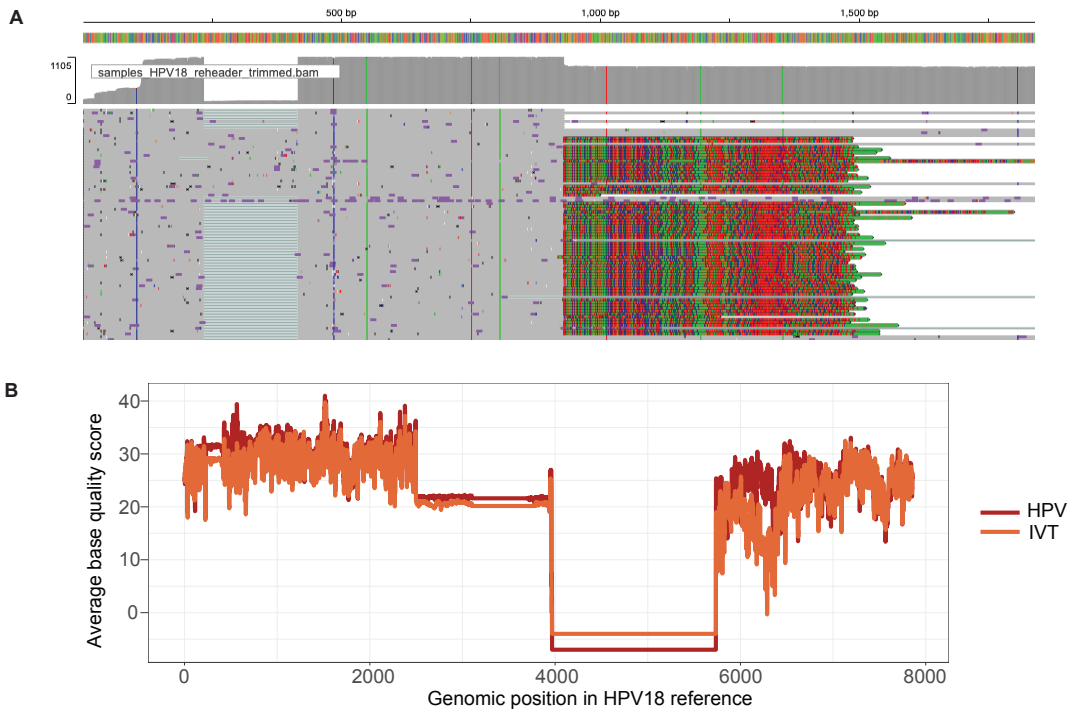

Figure S1. Quality check of HPV18-mapped reads using direct RNA-sequencing. (**A**) IGV screenshot illustrating the presence of softclipped (multi-coloured) regions in reads, derived from the human genome. (**B**) Average base quality per position for each of the HPV and IVT samples across the HPV18 reference genome. Zero quality indicates lack of coverage in these regions.

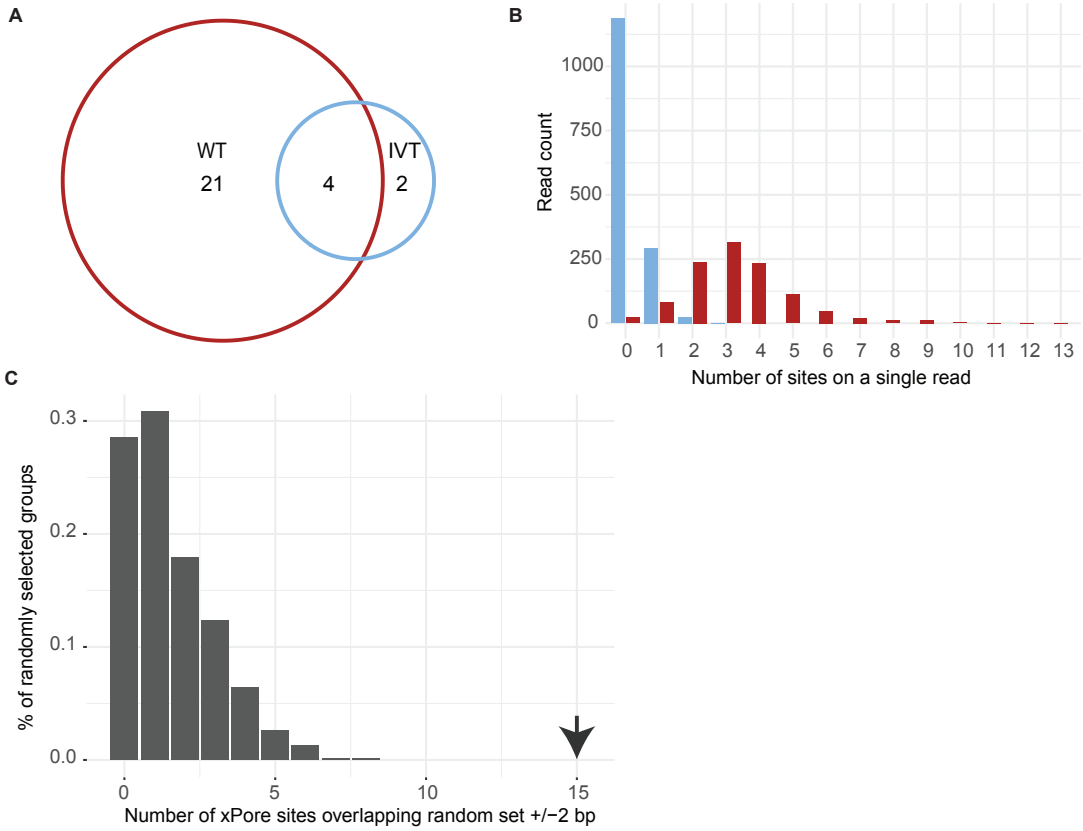

Figure S2. Properties of m^{6}A sites identified by Dorado in the WT and IVT sample.
(**A**) Venn diagram showing overlap between WT and IVT samples for Dorado detected sites. (**B**) Distribution of sites per molecule for the 25 WT-derived sites, for each of the WT and IVT samples. (**C**) Distribution of xPore sites overlapping sets of 25 randomly selected positions. Arrow represents the true overlap between the xPore significant sites and the Dorado detected sites (+/- 2bp).

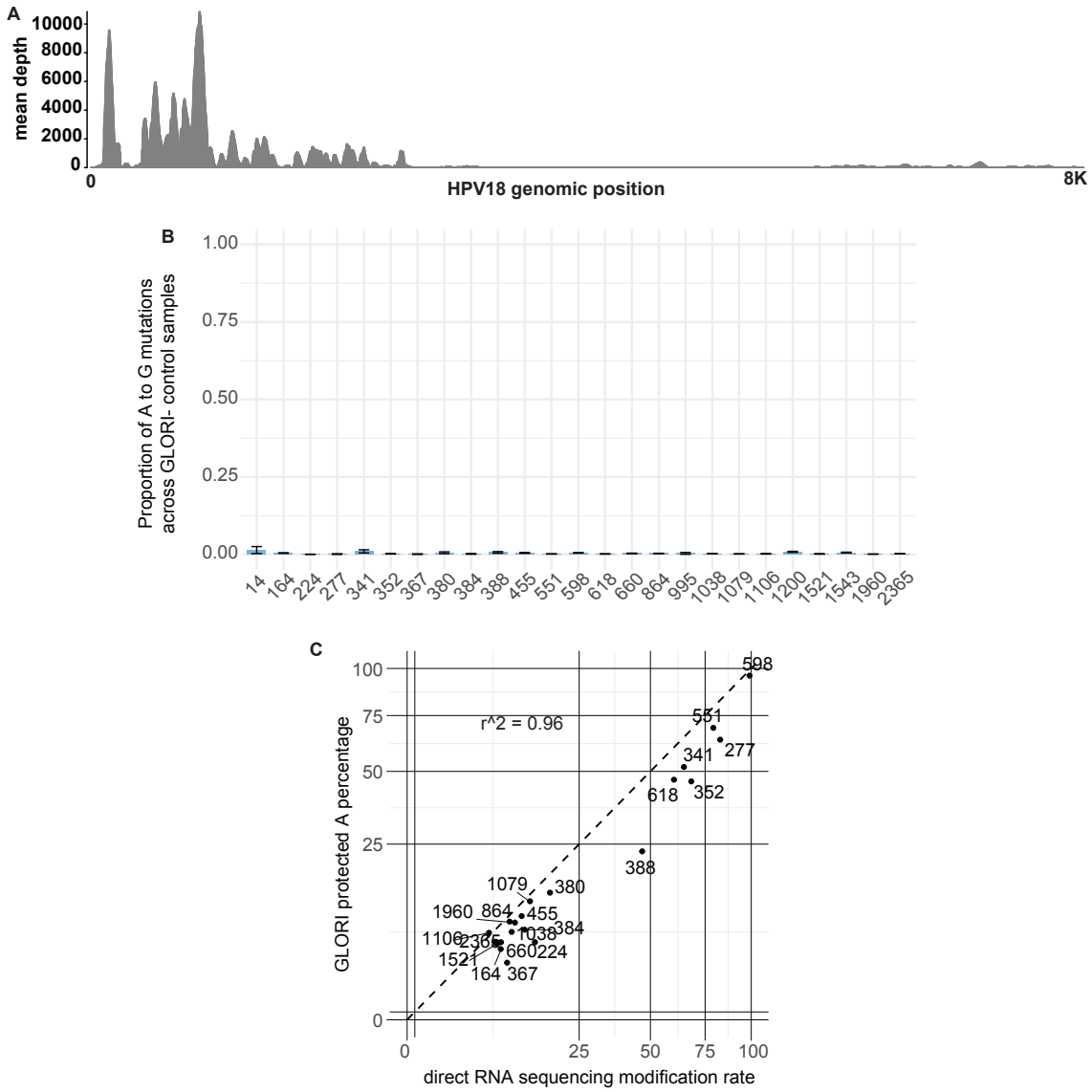

Figure S3. GLORI quality control analyses.
(**A**) Coverage of GLORI reads across the HPV18 genome, based on wild-type control samples for the knockdown, inhibition or hypoxia experiments (**B**) Proportions of A-to-G conversions based on the untreated GLORI (GLORI-) samples for the knockdown, inhibition or hypoxia experiments, for each of the 25 candidate m^6A sites identified by Dorado. (**C**) Scatter plot comparing detected modification rates from nanopore direct RNA-sequencing and GLORI for the 21 m^6A sites identified only in the WT sample.

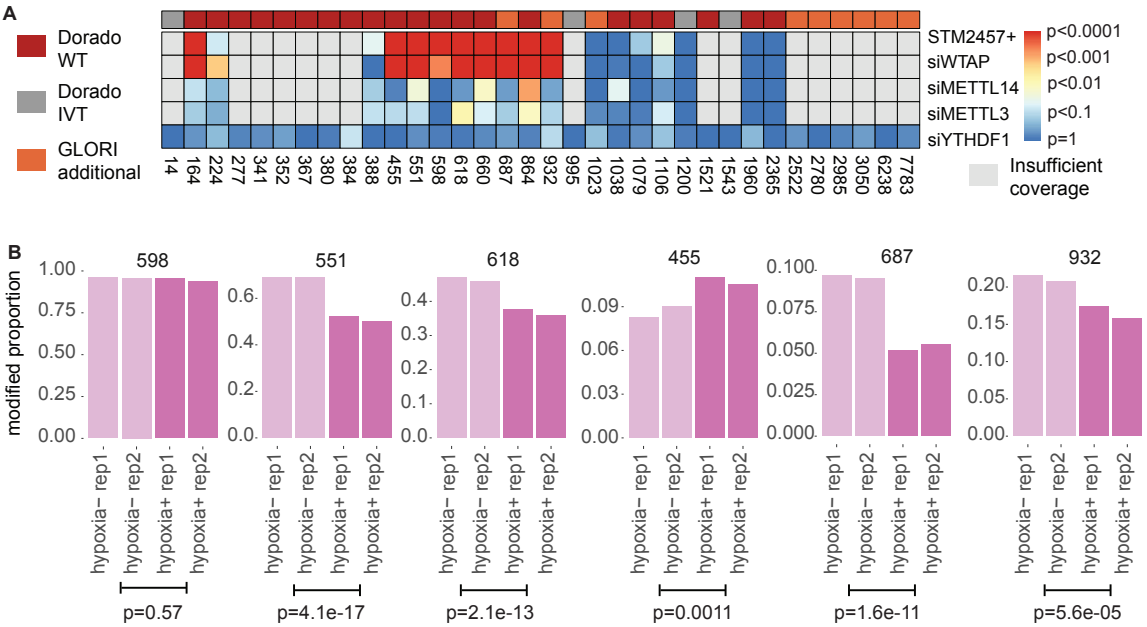

Figure S4. GLORI analysis of methylation changes based on knockdown, inhibition or hypoxia experiments.
(**A**) Heatmap showing significance of writer component knockdown or inhibition versus controls (siwriterControl) or of YTHDF2 knockdown versus controls (siYTHDF2control). Adjusted p-values from the binomial generalised linear model were converted into -log10(.) prior to plotting the heatmap. Coloured bar at the top represents the three groups indicated to the left of the plot. (**B**) Analysis of hypoxia samples. Bars represent proportion of protected As from treated GLORI samples, and p-values represent significance of hypoxia treatment based on a binomial generalised linear model.

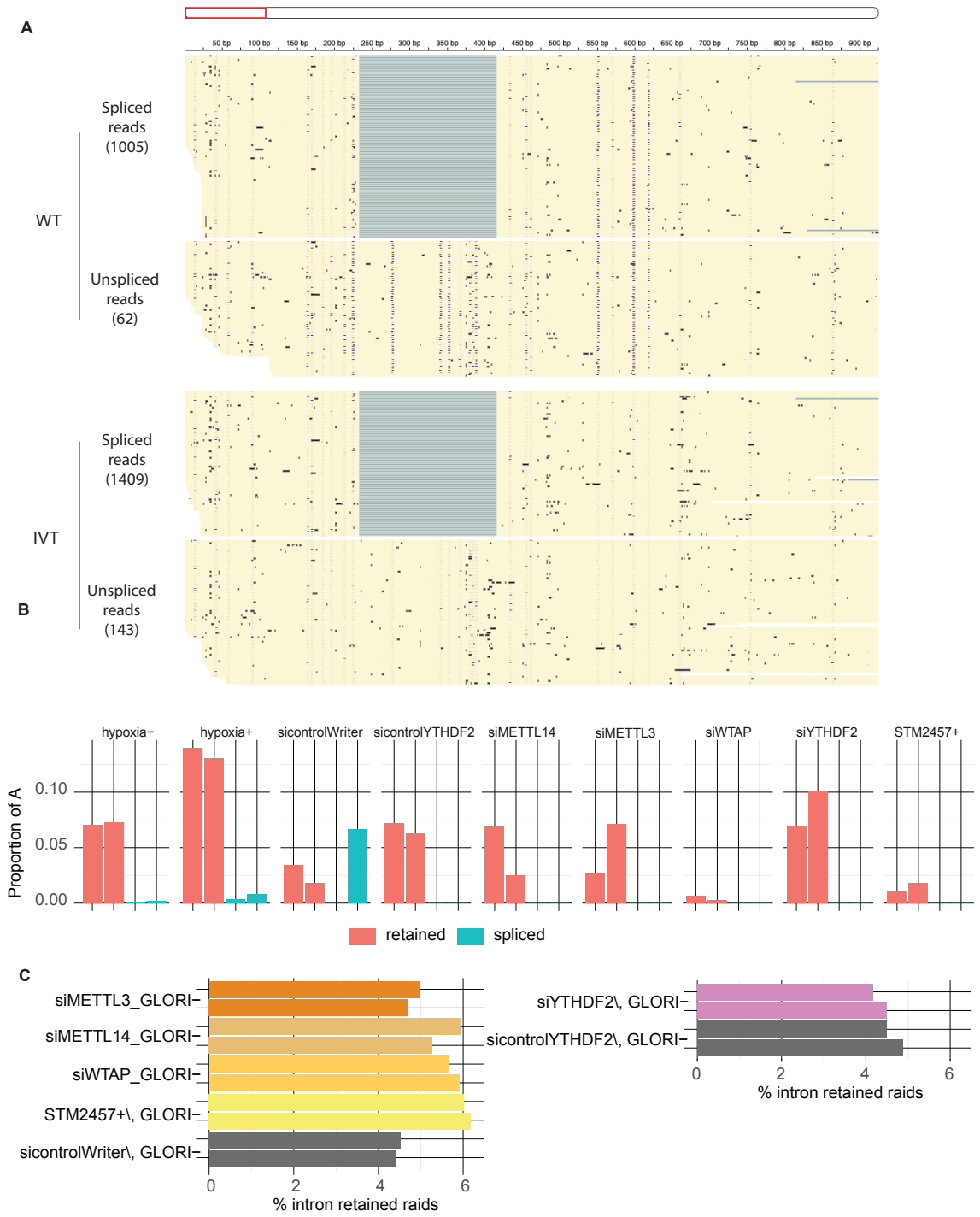

Figure S5. Analysis of E6-intron spliced or unspliced reads.
(**A**) Extended IGV view including spliced and unspliced reads from the IVT sample, blue dots represent m^6A methylation site confidence scores. The number of reads considered in each case is indicated on the left (**B**) Spliced versus unspliced protected A proportions for reads covering position 224 from the GLORI samples, where each plot shows the two replicates for the indicated condition. (**C**) Estimated percentage of reads retaining the E6 intron, per replicate for each of the writer knockdown or inhibition experiments, based on the untreated GLORI-samples. Grey bars represent the control (wild-type) case.

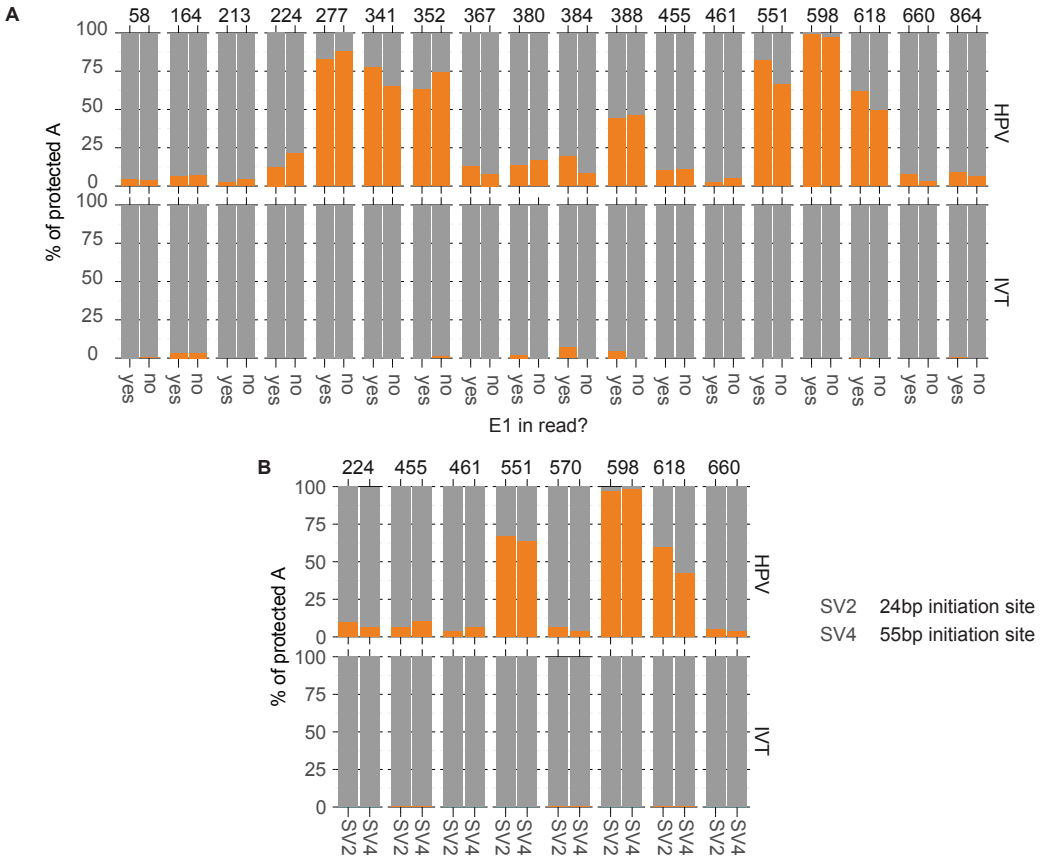

Figure S6. Analysis of differential methylation according to presence of E1 or use of transcription initiation site. (**A**) WT and IVT reads from direct RNA-sequencing split according to E1 presence, plotted is the modification rate inferred by modkit.(**B**) as (**A**) but reads for SV2 and SV4 are compared. All positions were tested using a Fisher's exact test (based on counts yes/no versus modified/unmodified reads and none were significant after multiple testing adjustments.

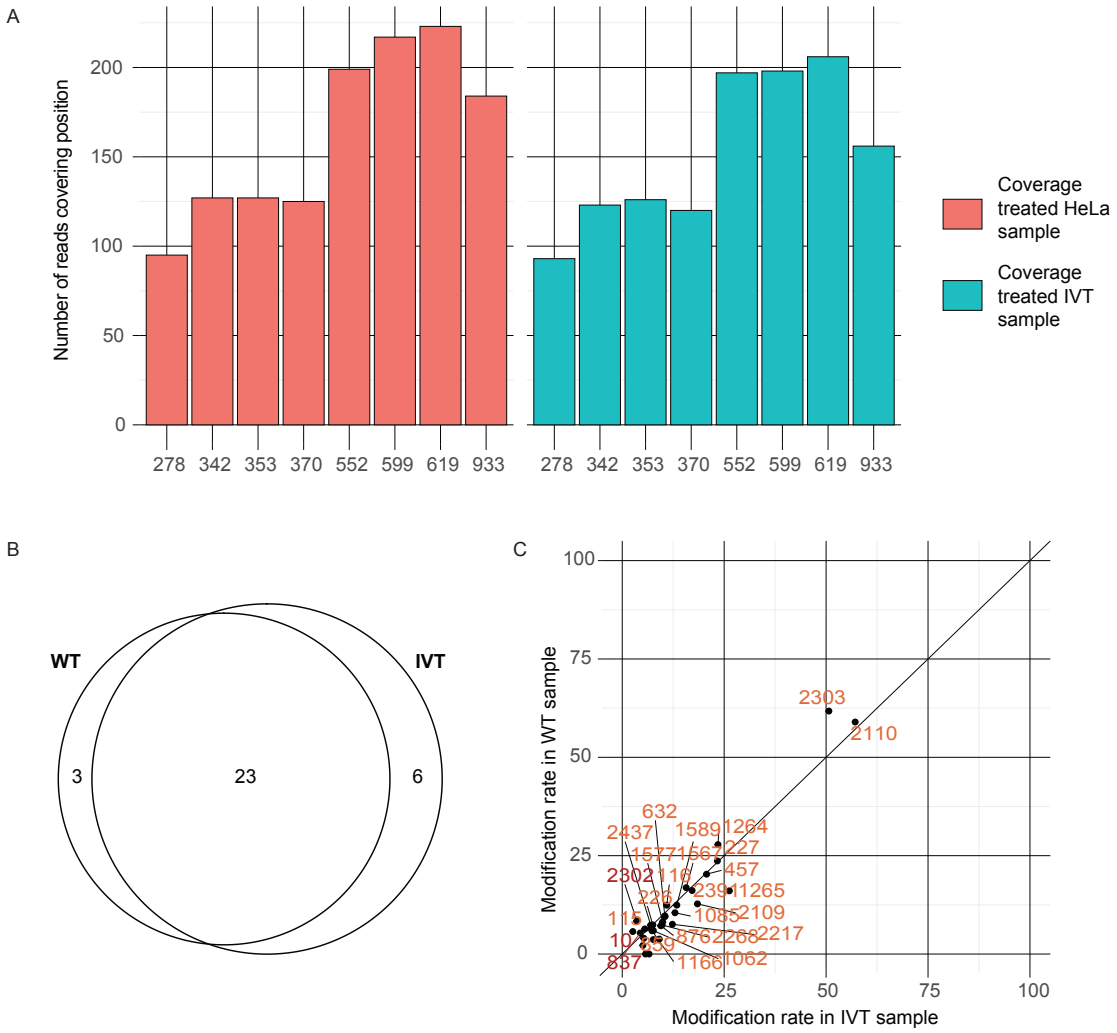

Figure S7. Analysis of m5C and Pseudouridine.
(**A**) Coverage of m5C-TAC-seq reads in chemically treated samples at the 8 candidate positions which were detected in the WT direct RNA-sequencing sample but not the IVT (**B**) Overlap between WT and IVT Psi sites identified by Dorado (**C**) Stoichiometry scatter plot for Psi.