A comprehensive survey of RNA modifications in a human transcriptome

Logan Mulroney[1,2,3*], Lucia Coscujuela Tarrero[1*], Paola Maragno[1], Carmela Rubolino[1], Simone Maestri[1], Mattia Furlan[1], Matteo Jacopo Marzi[1], Tom Fitzgerald[2], Tommaso Leonardi[1], & Mattia Pelizzola[1,4,†], Ewan Birney[2,†], Francesco Nicassio[1,†]

1 Center for Genomic Science of IIT@SEMM, Fondazione Istituto Italiano di Tecnologia, 20139 Milano, Italy

2 European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridgeshire CB10 1SD, UK

3 Epigenetics and Neurobiology Unit, European Molecular Biology Laboratory (EMBL), Monterotondo, RM 00015, Italy

4 Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milan, Italy

* These authors contributed equally
† Last and Corresponding authors

Mattia Pelizzola <Mattia.Pelizzola@iit.it>
Ewan Birney <birney@ebi.ac.uk>
Francesco Nicassio <Francesco.Nicassio@iit.it>

ABSTRACT

RNA modifications are critical for transcript function and regulation. Direct RNA nanopore sequencing offers a unique advantage by observing these modifications through characteristic alterations in ionic current signals. Various computational tools have been developed to detect RNA modifications, typically by comparing a sample of interest with a control lacking the target modification, often achieved through enzyme knockdown or knockout. We have developed a robust *in vitro* transcription protocol to generate a modification-free copy of any input RNA to detect all modifications that alter the raw ionic current signal relative to the canonical nucleotides during nanopore sequencing. We generated an *in vitro* transcribed sample from K562 cells and used Nanocompore to detect 26,619 modified RNA sites across 2,520 transcript isoforms from 1,742 genes. Of these, 56% are consistent with the well-characterized m6A modification. By inferring the identities of these modifications, we assessed differential usage and correlation patterns, revealing a significant co-occurrence between m6A and m5C modifications within the same transcripts. Additionally, some modifications were non-randomly associated with alternative splicing events. This study provides a comprehensive survey of RNA modifications across the transcriptome, demonstrating the utility of *in vitro* transcription coupled with direct RNA nanopore sequencing to simultaneously detect multiple modifications without the need for additional independent biochemical assays. The protocol is consistent in the future with more complex experimental designs, for example for differential modification usage between samples.

INTRODUCTION

The study of gene expression and regulation is essential to understand the mechanisms underlying cellular processes and diseases (Baylin and Jones 2011). Transcriptome analysis provides a comprehensive view of gene expression by detecting and quantifying RNA transcripts; depending on the details of the RNA preparation, one can also determine alternative splicing patterns. Cell metabolism is influenced at many levels (Esteller 2011), including transcription regulation (Henninger et al. 2021), splicing (Fu and Ares 2014), and translation (Gebauer and Hentze 2004). The biological function of RNA molecules is determined by their sequence, structure, and chemical modifications (Wan et al. 2014; Roundtree et al. 2017). RNA modifications are chemical alterations of nucleotides that often change the properties of RNA molecules, including their stability, structure, cellular localization and ultimately function (Roundtree et al. 2017; Gilbert et al. 2016; Delaunay et al. 2024). Despite the extensive and now routine use of RNA-seq experiments to assess RNA levels and alternative splicing, it has been challenging to assess RNA modifications in individual transcripts at single-nucleotide resolution. Although there is considerable evidence for the presence of RNA modifications and some understanding of their generic role, rarely are the RNA modifications resolved to individual transcripts or the role of individual RNA modifications on specific transcripts.

RNA modifications can be broadly classified into two categories: nucleoside modifications and base modifications. Nucleoside modifications involve the addition or removal of functional groups to the ribose or deoxyribose sugars of nucleotides, while base modifications involve the modification of the nitrogenous bases of nucleotides (Harcourt et al. 2017). The most abundant mammalian internal mRNA modification is N6-methyladenosine (m6A), which has been shown to regulate mRNA stability, splicing, and translation (Wang et al. 2014; Xiao et al. 2016; Wang et al. 2015). Another commonly identified RNA modification is 5-methylcytosine (m5C), which is found in tRNA, rRNA, and mRNA, where it is involved in translation and RNA folding (Schumann et al. 2020; Motorin and Helm 2010). Pseudouridine (Ψ) is the most

abundant mammalian RNA modification that affects RNA structure and stability (Kierzek et al. 2014), and 2'-O-methylation (Nm) is found in rRNA, snRNA, and mRNA, where it is involved in ribosome biogenesis and splicing (Monaco et al. 2018; Jady 2001). Methyl-Pseudouridine incorporation has been critical for the deployment of mRNA vaccines (Morais et al. 2021), allowing the rapid development of vaccines for infectious disease (Svitkin et al. 2022; Pardi and Krammer 2024) and the potential for individualized cancer vaccines. To date, approximately 170 naturally occurring RNA modifications have been identified (Boccaletto et al. 2018) and the community has expressed enthusiasm in systematically assessing their presence and relationships to other biological features (National Academies of Sciences, Engineering, and Medicine et al. 2024).

RNA modifications can be detected using a variety of methods, often through the results of chemical or enzymatic reactions that are specific for the RNA modification or modifications of interest (Kirpekar et al. 2000; Linder et al. 2015; Carlile et al. 2014). An emerging technique for detecting RNA modifications is Direct RNA Nanopore Sequencing (dRNA-seq) (Garalde et al. 2018). dRNA-seq processes RNA without any chemical intermediates, which doesn't erase the RNA modifications during sequencing (Workman et al. 2019). Thus, dRNA-seq has the potential to detect all RNA modifications which change the raw ionic current signal, without additional biochemical assays for particular RNA modifications. This detection happens in the context of individual molecules, thus it is possible to detect RNA modification co-occurrence and determine its relationship to the precise transcript structure.

Recently, there have been many studies that used dRNA-seq to investigate RNA modifications, some of which have developed new bioinformatic tools to detect the modified RNA bases (Furlan et al. 2021). Some of these tools can detect the RNA modifications from a single sample using characteristic ionic current shifts caused by the modification using training data of known modifications, while others require comparing the sample with a sample that is devoid of the modification(s) of interest (Furlan et al. 2021).

The methods that detect RNA modifications from a single sample typically use neural networks trained to detect particular RNA modifications (Furlan et al. 2021). These tools require laborious sample preparations of specific modifications to generate robust training data for the models. The two-sample comparison methods typically employ statistical tests at each position to find significant differences at each position between the two sample's distributions of ionic current features or systematic basecalling errors (Furlan et al. 2021). Single-nucleotide precision is challenging for these comparative methods because the ionic current for any state is derived from 5-9 nucleotides that occupy the sensitive region of the nanopore at each step of the motor regulated translocation (Manrao et al. 2012) and the RNA modifications can alter the ionic current for any of these positions.

Trained model methods generally outperform the two-sample comparative approaches for detecting the target RNA modification (Maestri et al. 2024). However, due to the challenges generating the training data, robust trained models are only available for a few well-characterized RNA modifications, such as m6A (Furlan et al. 2021). Comparative methods are thus invaluable when trained models are not available, or multiple RNA modifications are involved. These methods use a reference sample, that can be generated using two distinct strategies: altering the deposition or removal of a target RNA modification through genetics or environmental stimulation (Begik et al. 2021; Leger et al. 2021); or by generating a synthetic copy of the target RNA without any modified nucleotides (Tavakoli et al. 2023; Zhang et al. 2021). Although useful for synthesizing short oligonucleotide polymers, phosphoramidite chemistry is not practical for recreating an entire transcriptome (Song et al. 2021). Thus, to create a synthetic copy of a sample specific RNA transcriptome, *in vitro* transcription (IVT) is more practical.

Here, we present a robust method to generate substantial amounts of modification-free RNA transcriptome for direct RNA nanopore sequencing and the simultaneous detection of several

types of RNA modifications. We demonstrate the feasibility and utility of this approach by analyzing the epitranscriptome of a human cell line, K562, using Nanocompore, an analytical framework that finds sites significantly different between two sample by employing a Gaussian Mixture Model to cluster the ionic current in two dimensions for every read, position-by-position (Leger et al. 2021). We found 26619 modified positions, of which we could infer that 15003 (56.36%) sites were m6A and 4003 (15.04%) sites were m5C based on sequence motifs and annotations. We could detect multiple modifications in the same transcripts and in the same molecules. There were 1302 inferred m6A sites that had an inferred m5C site within 25 nt. In one transcript, KRTCAP2, we found a significant coordination of inferred m6A and m5C sites on the same molecules. We also found 448 genes with at least two isoforms that were differentially modified. This suggests that RNA modifications are coupled with alternative isoform usage, potentially influencing isoform-specific functions such as splicing, stability, and translation. Our IVT coupled with dRNA-seq approach provides a valuable tool for comprehensive RNA modification detection analysis. It can facilitate the study of multiple isoform-resolved RNA modifications simultaneously, without additional biochemical reactions for detecting specific modifications.

RESULTS

RNA in vitro transcription optimization

Recently, transcriptome wide IVT synthesis has been used with short-reads as a negative control for antibody based detection of m6A (Zhang et al. 2021). Like others (Tavakoli et al. 2023; McCormick et al. 2024), we thought this method could be adapted to long-reads to test for RNA modifications with dRNA-seq. The T7 RNA polymerase is commonly used for IVT reactions. To synthesize a sample-specific IVT transcriptome, the T7 promoter must be present on the 5′ end of a double stranded DNA template sequence for each desired transcript. This can be achieved by using a Template Switching Oligonucleotide that contains the T7 promoter sequence during cDNA synthesis with a template switching reverse transcriptase (**Figure 1A**). We tested several conditions for each step of the IVT RNA synthesis procedure

and found that the the NEB template switching kit combined with incubating the IVT reaction at 42°C, and in vitro polyadenylating the IVT RNA for 30 minutes were optimal conditions (see **Supplementary Results** and **Supplementary Figures 1 & 2** and **Supplementary Tables 1 & 2** for details on our optimization).

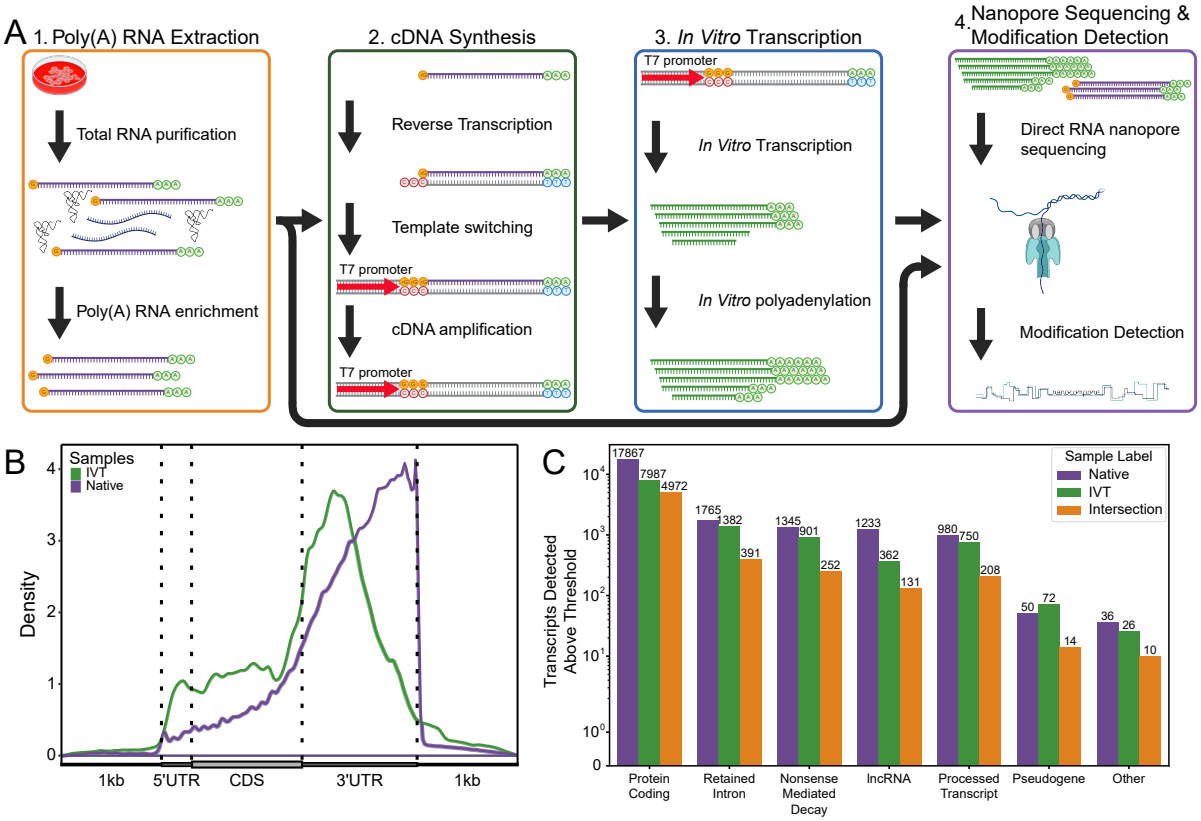

Figure 1: RNA *In Vitro* Transcription protocol and sequencing coverage: **(A)** Protocol diagram starting from RNA extraction and resulting in direct RNA nanopore sequencing of the native sample compared with the *in vitro* transcribed copy. **(B)** Metagene plot of the direct RNA nanopore sequencing coverage for the Native (purple) and IVT (green) samples across annotated mRNA transcripts. **(C)** Number of unique transcripts that were detected with more than 30 reads separated by biotype for the Native (purple), IVT (green), and by both the native and IVT samples (orange).

Using our optimized protocol, we obtained 5.4 and 2.9 million pass reads from two IVT samples sequenced on PromethION flowcells, resulting in 2.8 and 1.5 million forward primary alignments to the transcriptome reference, respectively (**Supplementary Table 3**). We observed a high rate of reverse alignments when aligned to the transcriptome reference which mostly initiated near short poly(C) sequences. From two Native samples, we obtained 9.1 and

8.7 million pass reads resulting in 7.7 and 7.3 million forward primary alignments to the transcriptome reference, respectively (**Supplementary Table 3**). As expected, the read coverage for the forward primary alignments were biased towards the 3′ end of most transcripts (**Figure 1B**). However, the IVT coverage was more even across the CDS and 5′ UTR compared to the Native coverage, likely due to the incomplete transcription products getting *in vitro* polyA tailed (**Figure 1B**).

The IVT reads do not need to perfectly match the length profile or detected abundance as the native dRNA-seq reads. Rather, the IVT reads need to cover enough of the transcriptome to provide the representative ionic current signal for each unmodified kmer in the sample. For example, Nanocompore requires 30 reads to cover each position by default (Leger et al. 2021). There were 4261 transcripts that had sufficient coverage for Nanocompore modification detection analysis in both the Native and IVT samples (**Figure 1C**).

RNA modification detection

Nanocompore, a tool we developed, is a comparative tool that was recently reported to achieve good performances, especially in case of high-coverage samples (Maestri et al. 2024). Moreover, it allows extracting modification probabilities at base pair level resolution and read-level, a feature that is highly desirable for assessing the co-occurrence of multiple modifications on the same RNA molecule (Leger et al. 2021).

First, we wanted to test the extent to which the differences between the hardware and software for the MinION and PromethION platforms altered the ionic current signal. We tested this by using Nanocompore with the PromethION IVT samples and either the MinION or PromethION Native samples. Indeed, we found that the modifications detected were different depending on which platform was used for the Native samples (**Supplementary Figure 4A & C**). These differences remained even after balancing the coverage between the two platforms

(**Supplementary Figure 4B & D**). We therefore decided to use only the PromethION data for both IVT and Native samples for the rest of our tests.

From the 4261 transcripts, there were 2.97M total positions analyzed by Nanocompore (see methods and **Supplementary table 4**). We found 26619 significant positions in 2520 transcripts from 1742 genes (**Supplementary table 4**). The number of significant sites per transcript ranged from 1 to 149, with the largest population of transcripts containing only a single modified position (

A).

The IVT sample is composed entirely of canonical nucleotides, thus all types of RNA modifications are potentially identified. However, from this analysis alone, the identity of these modifications is unknown. The RNA modification identity can be inferred by matching the sequence context with known motifs for RNA modification writing enzymes or by comparing the sites identified with Nanocompore to RNA modification annotations, such as those compiled by RMBase3.0 (Xuan et al. 2024). We compared the distribution of the Nanocompore significant sites with those annotated by RMBase across a metagene representation (see methods), and found that both the Nanocompore sites and RMBase annotations have a peak near the stop codon (**Figure 2B**). It's been previously documented that m6A is predominantly detected in the 3′ UTR adjacent to the stop codon (Meyer et al. 2012), as can be seen when only the m6A annotations are plotted in the metagene (**Figure 2B**). The Nanocompore sites had more density in the CDS than what is annotated by RMBase. Considering that m6A is the dominant modification annotated in RMBase and Nanocompore can detect many other modifications which may be distributed differently across the transcripts, this suggests that the IVT-based analysis identified sites of RNA modifications that are different from m6A.

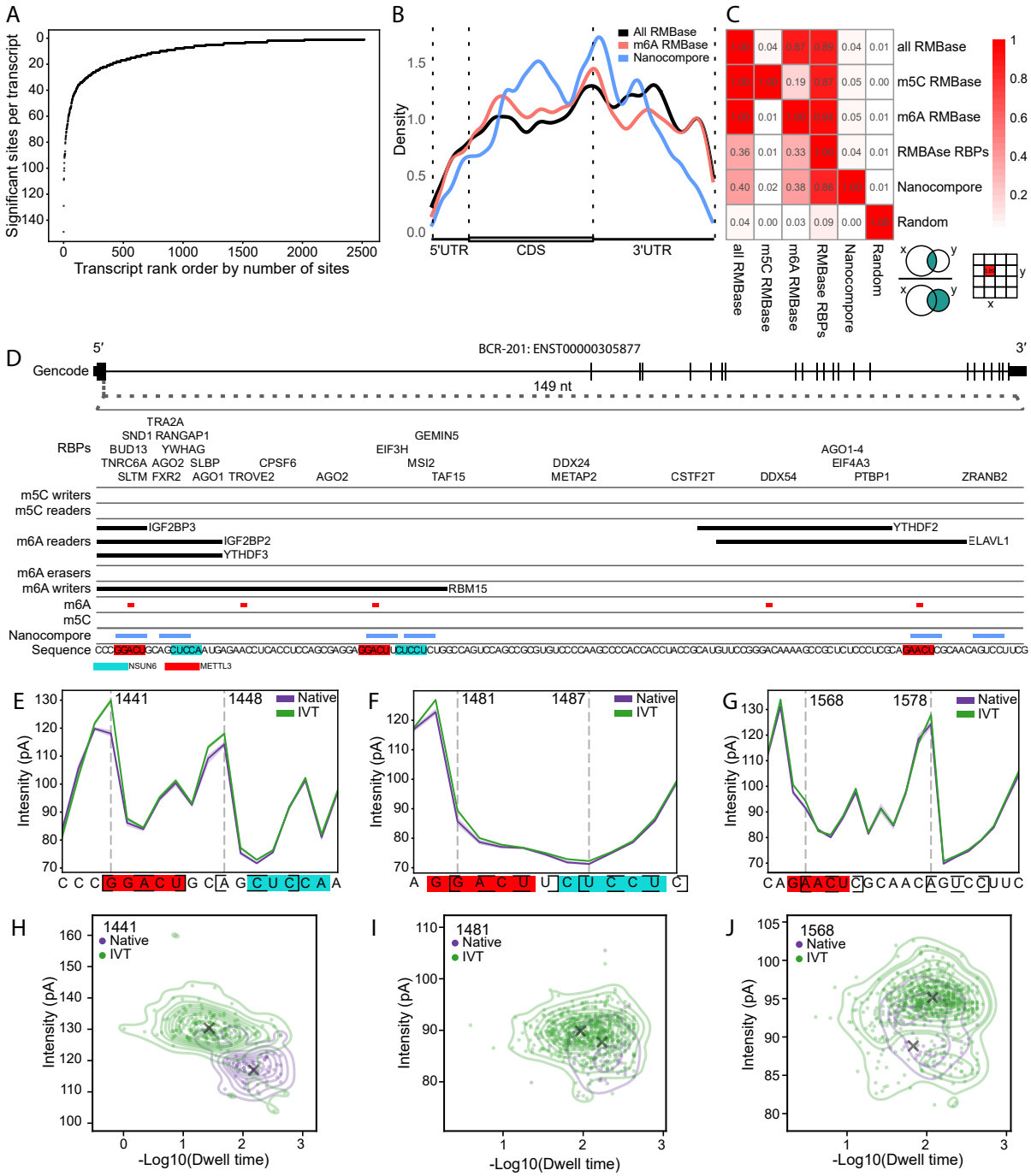
A
Significant sites per transcript
Transcript rank order by number of sites
B
All RMBase
m6A RMBase
Nanocompore
Density
5'UTR
CDS
3'UTR
C
all RMBase
m5C RMBase
m6A RMBase
RMBase RBPs
Nanocompore
Random
D
Gencode
BCR-201: ENST00000305877
149 nt
RBPs
m5C writers
m5C readers
m6A readers
m6A erasers
m6A writers
m6A
m5C
Nanocompore
Sequence
NSUN6
METTL3
E
F
G
Native
IVT
Intesnity (pA)
H
I
J
-Log10(Dwell time)

Figure 2: Detected RNA modifications. **(A)** A scatterplot of the number of Nanocompore significant sites per transcript detected in rank order. Each point is a transcript isoform, the y-axis is the number of Nanocompore significant sites detected for that isoform, the x-axis is the rank order by number of Nanocompore Significant sites. **(B)** A metagene plot of the Nanocompore significant sites (blue), the m6A sites annotated by RMBase3.0 (red), and all the modifications annotated by RMBase3.0 (black). The vertical dotted lines denote the borders for each major region of the meta mRNA transcript (5′ UTR, CDS, and 3′ UTR). **(C)** Heatmap reporting the overlap of the Nanocompore significant sites with different features annotated in RMBase3.0. The value in each cell of the matrix is calculated by dividing the intersection between the positive row and column bins divided by the total number of positive row bins. See diagram for reading the values in each cell. **(D)** A 149 nt wide region IGV trace of an example transcript of BCR, ENST00000305877. Six of the nine detected Nanocompore significant sites for this transcript are found within this region. The RMBase3.0 annotations are displayed in the first 8 rows. The RBPs row is defined as the binding sites identified for RNA modification associated RNA binding proteins that are not explicitly denoted by other classifications. The m5C and m6A effector rows are the binding sites annotated for those enzymes, and the m6A and m5C rows are defined as the annotated sites for m6A and m5C. The Nanocompore row is the Nanocompore significant sites, and the sequence is the transcriptomic sequence for the region. The sequence regions highlighted in red match m6A writer enzyme METTL3 known sequence recognition motif, DRACH. The sequence highlighted in cyan is the known m5C writer enzyme NSUN6, CUCCA and a close match CUCCU. **(E)** Line plot of the mean ionic current intensity (y axis) for the Native reads (purple) and the IVT reads (green) for the region surrounding the first pair of Nanocompore significant sites. The line is the mean for all reads aligned to that position and the highlighted region is one standard deviation. The vertical dotted line represents the Nanocompore significant sites, 1441 and 1481. The x-axis is the sequence for the region. The color highlights are the same as in **(D)**, and the dashed box represents the kmer that occupies the sensitive region of the nanopore for the significant position. **(F)** and **(G)** are line plots for the following two regions. **(H)** Two dimensional distribution of the ionic current intensity (y-axis) and the -log10 of the dwell time for each read at position 1441. The Native reads are purple and the IVT reads are green. The GMM centers are marked with black X. **(I)** and **(J)** are Two dimensional distributions of the ionic current intensity for positions 1481 and 1568 respectively.

To more precisely evaluate the overlap between the annotations and the Nanocompore significant sites, we binned the genome into 25 nucleotide wide non-overlapping windows and looked at the bins that were shared between the two sets (see methods, **Figure 2C** & **Supplementary figure 3**). We found that 38% of the bins with Nanocompore significant sites overlapped with a bin containing an RMBase m6A site. When the Nanocompore positive bins were compared to all RNA modifications annotated by RMBase, there were only an additional 2% of the bins that overlapped (40% overlap in total). However, 89% of the Nanocompore positive bins overlapped with an annotated RNA binding protein associated with modifications. This represents a 9-fold increase in overlaps when compared to randomly shuffling the

Nanocompore significant sites (**Figure 2C**). The data and analysis support using IVT baselines for an unbiased calling of the epitranscriptome.

Coordination between RNA modifications

To understand the possible RNA modifications identified by Nanocompore, we examined the sites found in BCR, a gene typically found mutated in leukemia, the source tissue for K562 cells (Chandrasekhar et al. 2019). We found nine Nanocompore significant sites in BRC, six of which were found in the first exon (**Figure 2D**). Surprisingly, these six significant sites appeared in pairs, where the first site of each pair overlapped with both m6A sites annotated by RMBase and the DRACH motif (D=[AUG], R=[AG], A, C, H=[AUC]), the main sequence recognition motif for the m6A writer enzyme, METTL3 (Dominissini et al. 2012) (**Figure 2D**). A site paired with an apparent m6A site overlaps with the m5C writer enzyme, NSUN6, sequence recognition motif, CUCCA (Selmi et al. 2021). The second site overlaps with a sequence that closely resembles the NSUN6 motif, CUCCU. The third paired Nanocompore significant site didn't closely resemble any known RNA modification writing sequence recognition motif. All six of these sites have detectable ionic current differences between the Native and IVT samples (**Figure 2E-2G**). In addition, they also separate into two distinct clusters in terms of ionic current and dwell time (Nanocompore GMM correct p-value < 0.01 and absolute value of the Log Odds ratio > 0.5, **Figure 2H-2J**).

This proximity of several Nanocompore significant sites suggests that there may be an association between significant sites on the same RNA molecules. To test this, we determined the modification status for each read at every Nanocompore significant site by using the probability of being assigned to the modified cluster in the GMM. Then we performed a chi-squared test on the counts of patterns of modifications for each transcript isoform, under the assumption of independence of each modification (see methods). Only two transcripts were significant at FDR < 0.01% for Nanocompore site coordination, YDJC and KRTCAP2. Both transcripts had three Nanocompore significant sites, of which we were able to infer the

modification identity for all the sites in KRTCAP2 (**Figure 3A-C**). Specifically, the Nanocompore site I overlaps with the m6A writer enzyme, METTL16 sequence recognition motif, UACAG (Mikutis et al. 2020), and site II is adjacent to the NSUN6 sequence recognition motif, CUCCA (Selmi et al. 2021). Site III overlaps with a RACH motif, an RMBase annotation for m6A, and an iCLIP signal for WTAP, a protein that forms a complex with METTL3 and METTL14 (Ping et al. 2014) (**Figure 3A-3C**). Of the eight possible permutations of single-molecule modification status for the three sites, one didn't have enough molecules in the class to pass our filtering requirements (see methods) and two were significant with a chi-squared test (**Figure 3D**). The observed counts were significantly higher than the expected counts when the Nanocompore sites I & II were both modified on the same molecules, and significantly lower than the expected counts when the sites I was unmodified and site II was modified on the same molecules (**Figure 3D & 3E**).

It was surprising to observe two examples (BCR and KRTCAP2) with apparent m6A sites in proximity to apparent m5C sites. However, this is consistent with the top ten enriched Nanocompore kmers (**Figure 3F**). Three were DRACH motifs and four were associated with sequences enriched in NSUN6 miCLIP experiments (Selmi et al. 2021) (**Figure 3F**). Given this enrichment for apparent m6A and m5C sites, we set to broadly infer which Nanocompore significant sites were m6A and m5C in the transcriptome (see methods). We could infer that 15003 (56.36%) significant sites were m6A, based on the proximity to a DRACH (METTL3), UACAG (METTL16) sequence motif, or an annotated m6A site by RMBase3.0 (Xuan et al. 2024). We also identified 4003 sites as m5C using the NSUN6 sequence motif (CUCCW) or an annotated m5C site in RMBase3.0. There were 11298 inferred m6A sites that co-occurred on the same transcripts as 3831 inferred m5C sites. Using these inferred m6A and m5C sites, we determined the distances from m6A sites to the closest m5C sites. We found the peak distance from inferred m6A sites to the closest inferred m5C sites was 23.07 nt. There were 1302 (11.52%) distances from inferred m6A sites to the closest inferred m5C sites that were within 25 nt. In contrast, the peak of the randomly shuffled m6A site distances distribution was

243.47 nt and only 77 (1.19%) were within 25 nt of an inferred m5C. The proximity of the bulk inferred m6A and inferred m5C sites suggests that there is some shared function of these two RNA modifications. However, the lack of signal in the single-molecule analysis of RNA modification coordination on the same transcripts suggests that these are random. More likely, we lack the power to adequately detect coordination of RNA modifications on single-molecules with the number of reads we have for each transcript.

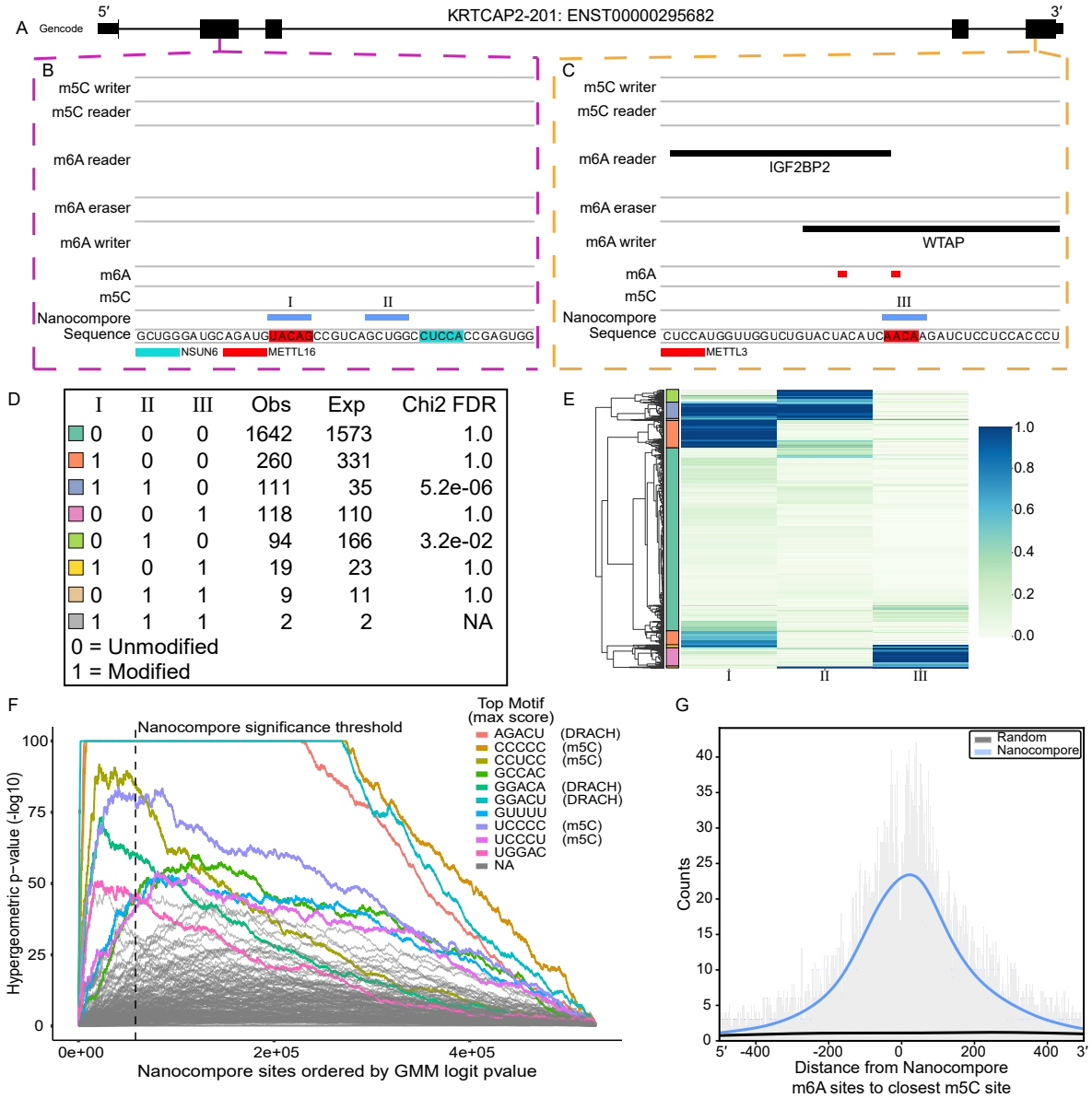

Figure 3: RNA modification co-occurance. **(A)** Gencode v36 gene model for KATCAP2 isoform ENST00000295682. **(B)** Expanded IGV region of exon two showing the RMBase3.0 annotations for m5C writer and reader enzymes, m6A reader, writer, and eraser enzymes, annotated m6A and m5C sites, the Nanocompore significant sites, and the transcriptomic sequence for the region. The two Nanocompore significant sites are labeled I and II. The sequence recognition motifs for the m6A writer METTL16 (red) and the m5C writer NSUN6 (cyan) are highlighted in the sequence. **(C)** The same as in **(B)**, but for a region of exon five and the sequence recognition motif for the m6A writer METTL3 (red) is highlighted in the sequence overlapping with site III. **(D)** Table showing the different possible permutations of modifications at the three nanocompore significant sites, I, II, & III on each sequenced transcript, the observed and expected counts for each permutation, and the FDR multiple testing corrected chi2 pvalue for each permutation. A 0 indicates that the site is unmodified for that permutation, and a 1 indicates the site is modified (>= 75% chance of belonging to the modified GMM cluster). The colored boxes correspond to the groups in panel **(E)**. **(E)** Hierarchical clustering of the molecularly phased RNA modifications for sites I, II, & III. The heat is the probability that the read belongs to the modified cluster. The colored boxes correspond to the permutations in panel **(D)**. **(F)** Sylmar analysis for the kmers with GMM corrected pvalues between 0 and 0.5. The kmers are plotted in rank order according to their GMM pvalue along the x-axis and the y-axis is the -log10 of the hypergeometric pvalue. The top ten max scoring kmers are highlighted and the kmer sequence is indicated in the legend. If the kmer's association with different RNA modifications is indicated in parentheticals next to the kmer. All other kmers are colored in grey (NA). **(G)** Histogram of distances between inferred Nanocompore m6A sites and the closest Nanocompore inferred m5C site on the same transcript isoform. Zero is the inferred m6A site, negative distance indicates the inferred m5C is upstream of the inferred m6A and positive distance indicates that the inferred m5C is downstream of the inferred m6A. The Nanocompore sites are in blue, and randomly shuffled sites are in black.

Differentially Modified Isoforms

We identified 1742 genes with at least one Nanocompore significant site. Among these, 1294 genes only had one isoform and 448 genes had multiple isoforms (**Figure 4A**). As expected, most isoforms only had a single significant site (**Figure 2A**). However, surprisingly, these sites often differed across isoforms of the same gene (**Figure 4B**). Approximately 70% of Nanocompore significant sites are detected in a single isoform. Interestingly, in the case of PABPC1, a gene that binds poly(A) tails and is involved with translation and NMD (Peixeiro et al. 2012), we detected a single position that was significant in 14 different isoforms (**Figure**

4B). We were unable to identify the possible RNA modification at this site. The significant kmer sequence is UUUUG, and there are no annotated RNA modifications in proximity of the site.

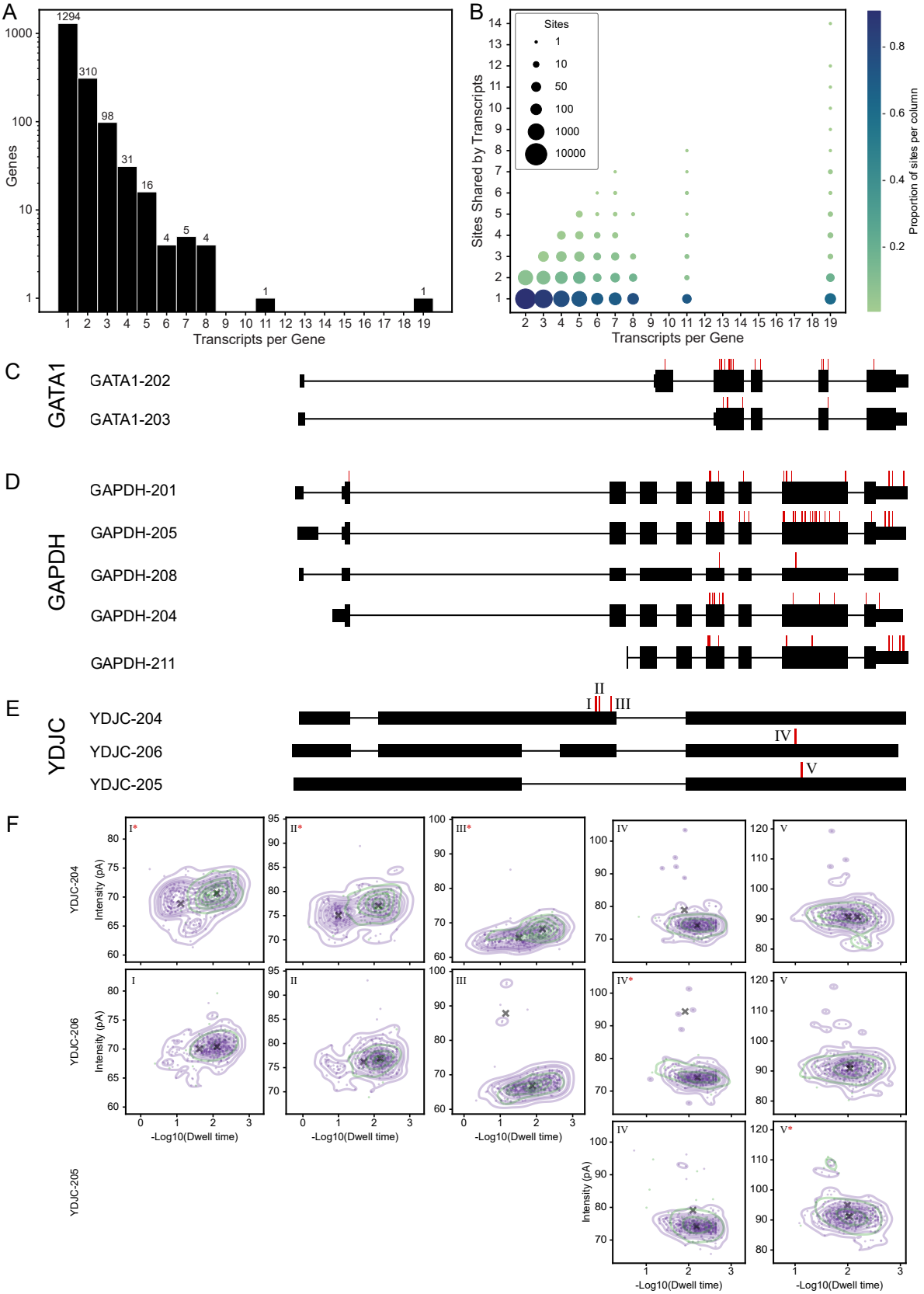

Figure 4: Isoform differential RNA modifications. **(A)** Histogram of the number of isoforms per gene detected with at least one Nanocompore significant site. **(B)** Dot matrix plot for the number of Nanocompore significant sites that are found in common with multiple isoforms. Dot area represents the number of sites, and the color indicates the proportion of sites for a column in the dot matrix plot. **(C-E)** Example gencode v36 transcript isoform models detected with at least one Nanocompore significant site for **(C)** GATA1, **(D)** GAPDH, and **(E)** YDJC. All transcript models are in the 5′-to-3′ orientation. Red lines indicate a Nanocompore significant site. **(E)** has the five Nanocompore significant sites labeled I, II, III, IV, and V, which correspond to the panels in **(F)**. **(F)** Ionic current distributions for the five Nanocompore significant sites found for YDJC. The y-axis is the ionic current intensity, and the x-axis is the -log10 of the dwell time. Reads from the Native sample (purple) and IVT sample (green) are shown. Black X's indicate the GMM cluster centroids. The five sites are labeled in the upper left corner and a red astritic indicates that site was significant (Adjusted GMM logit pvalue <= 0.01 and Absolute value of the log odds ratio > 0.5).

The landscape of Nanocompore significant sites across different isoforms is complex, as can be seen in three representative genes, GATA1, GAPDH, and YDJC. In agreement with our observations across all isoforms, most of the Nanocompore significant sites are found in one of the isoforms, with few, if any, shared among the isoforms (**Figure 4C-4E**). A closer inspection of the ionic current data for YDJC isoforms shows that there is clear separation between the native and IVT reads for YDJC-204(top) at positions I, II, and III. YDJC-206 (middle), is not significant for a modification at these three positions and as expected, there is more overlap in the ionic current between the Native and IVT samples. Isoform YDJC-205 (bottom) is lacking the exon for positions I, II, and III, and thus doesn't have any ionic current alignments for those positions (**Figure 4F**). These results demonstrate the complexity of how RNA modifications are distributed throughout the transcriptome and allows for further investigations into isoform differential modifications and possible functional differences.

DISCUSSION

In this study, we tested RNA *in vitro* transcription (IVT) to generate a modification-free transcriptome as a reference sample to detect RNA modifications using dRNA-seq. In our

study, the IVT was limiting the analysis to a subset of the transcripts detected by the Native sample (**Figure 1C**). We can envision overcoming this limitation by either sequencing the IVT sample more deeply, or by altering the modification comparison framework to use the IVT data independent of the positions where the strands align but instead a pool of ionic current data akin to an ionic current kmer look-up table.

Despite the limitations of the IVT sample, we were still able to detect 26619 positions that were significant when compared with the Native sample (GMM logit FDR p-value <= 0.01 and the absolute value of the Log Odds Ratio > 0.5) (**Supplementary Table 4**). Since the IVT sample is RNA modification-free, all RNA modifications are detected simultaneously, however, it is not possible to know the identity of the detected RNA modification from this analysis alone. As m6A is the most abundant internal mammalian mRNA modification (Anreiter et al. 2021), it was unsurprising that the Nanocompore significant sites we detected were distributed similarly to previously annotated m6A sites (Meyer et al. 2012), with some differences (**Figure 2B**). We could infer that 56.36% significant sites were m6A and that 15.04% sites were m5C using known m6A and m5C writer sequence motifs and annotated m6A and m5C sites from RMBase3.0 (Xuan et al. 2024).

Surprisingly, 11298 of the inferred m6A sites co-occurred on the same reference transcripts as 3831 of the inferred m5C sites, with 1302 (11.52%) of the distances between them within 25 nt of each other (**Figure 3G**). We also found significant enrichment for proximal inferred m6A and m5C on the same molecules for KRTCAP2 (**Figure 3**). Recent advances in m5C detection methods are also finding that m5C is more prevalent in mRNA than previously thought (Hartstock et al. 2023; Lu et al. 2024). Furthermore, Mateos et. al. also found that m6A and m5C are in close proximity and co-occurring on the same molecules at higher than expected rates in poly(A) RNA samples from HEK293 and HeLa cells using a different nanopore RNA modification detection method (Acera Mateos et al. 2024).

It has been previously documented that m6A and m5C jointly promote methylation of each other, and together enhance protein translation of CDKN1A in response to stress agents (Li et al. 2017). Although the mechanism for how m6A promotes m5C methylation and vice versa is not entirely clear, it is becoming apparent that the relationship between m6A and m5C is more widespread than previously thought as better m6A and m5C detection methods become available. Furthermore, there may be temporal components to the mechanism of methylation cooperation. There is a precedent for RNA modifications being deposited in successive order in rRNA biogenesis (Sloan et al. 2017). However, it is unclear if the order of modification deposition is step-wise or stochastic at each stage of rRNA maturation (Sloan et al. 2017). Simultaneous detection of different RNA modifications on single-molecules could help resolve potential temporal components of RNA modification deposition cooperation.

Not only can long-reads from dRNA-seq be used to molecularly phase multiple RNA modifications, but also map efficiently to isoforms (Pardo-Palacios et al. 2024). This can be used to resolve isoform specific modification patterns (Gleeson et al. 2024). In this study, we found 1742 genes with Nanocompore significant sites, among which, 448 had significant sites in at least two different isoforms (**Figure 4**). The number of isoforms per gene that we found with Nanocompore significant sites ranged from two (310) to nineteen (1) Most of the sites were found in only one of the isoforms indicating that modification deposition was not consistent across isoforms. In some cases there is a hyper-modified isoform, and a hypo-modified isoform like seen in GATA1 In other cases, there was a mix of shared and unique significant sites, like seen in GAPDH (**Figure 4C-D**). And in a few cases, there were Nanocompore significant sites that were all unique to the different isoforms as seen in YDJC (**Figure 4E**).

There are several possible technical and biological reasons for detecting different patterns of RNA modifications in the different isoforms. One technical complication is the number of reads mapped to each isoform. The GATA1 isoform, GATA1-202 (**Figure 4C top**), has more Nanocompore significant sites, but is also the dominant isoform detected with 10-fold more native reads than GATA1-203 (**Figure 4C bottom**), 3613 and 342 native reads respectively. For detection by Nanocompore, more than 2000 reads are suggested to detect a modification with 10% stoichiometry (Leger et al. 2021). Additionally, the IVT reads are not always evenly distributed between the various isoforms and thus certain sites in different isoforms may not be detectable due to a lack of IVT reads at that site.

One possible biological reason is the exon structure of the isoform. For example, the three YDJC isoforms we detected all have 470, 584, and 446 native reads respectively (**Figure 4E)**. However, YDJC-204 is the only isoform significant for a modification at sites I, II, and III (**Figure 4E top)**. Unsurprisingly, YDJC-205 (**Figure 4E bottom**) has no reads that cover these sites because it lacks the exon. YDJC-206 (**Figure 4E middle**) does have the exon, but was not significant at any of the three sites. Uzonyi et. al. recently proposed a model for m6A deposition, where the splicing machinery excludes the methylation machinery from accessing potential target sites for modification deposition (Uzonyi et al. 2023). The only structural differences between YDJC-204 and YDJC-206 is an alternative TSS and an intron upstream of sites I, II, and III. The YDJC-206 upstream intron could be preventing the modification deposition. The ionic current distributions of the Native reads for YDJC-204 and YDJC-206 separate into two distinct clusters at the three sites whereas the IVT reads form a single cluster (**Figure 4F**). Despite the few possibly modified reads for YDJC-206, they are greatly reduced when compared to YDJC-204 and do not pass the significance threshold for our Nanocompore analysis. Based on visual inspection of the ionic current distributions, it is less clear why YDJC-206 is significant at site IV and YDJC-205 is significant at site V. It is possible that these two sites are false positives. Alternatively, the underlying RNA modification could have a more subtle effect on the ionic current that is detected by the GMM and not clear from visual

inspection. Without a clear indication of which RNA modification is present at these two sites, it is difficult to resolve the truth at these two sites without a comprehensive RNA modification writer enzyme KD/KO screen.

All together, these results demonstrate the utility and challenges using an IVT transcriptome for detecting the positions of RNA modifications with dRNA-seq. Identifying different RNA modifications on the same molecules gives insights into possible coordination of different RNA modifications, and their distribution amongst different isoforms. Detecting these coordinated RNA modifications can be used to design follow-up experiments to determine the mechanism or function of these coordinated sites.

Furthermore, evaluating single-sample trained RNA modification detection models remains a challenging problem. As these methods become more widely adopted, repositories of IVT dRNA-seq datasets are invaluable true negative test cases to evaluate false positive rates. It is also unclear how quickly robust trained models can be produced for the nearly 170 known naturally occurring RNA modifications (Boccaletto et al. 2018) and the nearly limitless possible synthetic RNA modifications (Stephenson et al. 2022). Until that time, comparing native dRNA-seq datasets to matched IVT dRNA-seq datasets permits detecting the multitude of known RNA modifications, and possibly highlighting possible locations to investigate further for potentially novel RNA modifications that wouldn't be detected by these trained models.

In conclusion, IVT-based calling of RNA modifications currently represents the ideal strategy for deciphering the location of the multitude of modifications that make up the whole epitranscriptome, whose composition and dynamics remain to be investigated.

METHODS

Cell culture Total RNA extraction and mRNA purification

K562 cells were grown in RPMI-1640 medium supplemented with 10% FBS (Sigma-. Aldrich F2442). Total RNA was extracted from 10 Million cells using Qiazol (Qiagen 79306) and RNeasy Micro Kit (Qiagen 74004). mRNA purification was performed with 100ug of Total RNA using µMACS™ mRNA Isolation Kit (Miltenyi Biotec 130-075-201) following the manufacturer's protocol.

cDNA synthesis protocol

Reactions for cDNA synthesis were prepared in 0.2 ml nuclease free PCR tubes on ice. First, 100 ng of mRNA were hybridized with 1 µM of the PolyT primer (5′-T_{30}VN-3′) and 1 mM dNTPs in 6 µl. The samples were mixed gently by pipetting and centrifuged briefly. Then Incubated for 5 minutes at 70°C and left at 4°C until next use. Second, the hybridized RNA was mixed with 4 µl of RT buffer (1X Template Switching buffer, 3.75 µM Template Switching Oligo (5′-ACTCTAATACGACTCACTATAGGGAGAGGGCrGrGrG-3′), 1X Template switching RT Enzyme), and incubated at 42°C for 90 minutes, heat inactivated at 85°C for 5 minutes. The 10 µl of the first strand RT products were PCR amplified in 50 µl total volume of PCR buffer (1X Q5 Hot Start Master Mix (NEB #M0466S) and 0.2 µM T7 extension primer (5′-GCTCTAATACGACTCACTATAGG-3′)), with 15 cycles of 98°C for 10 sec, 68°C for 15 sec, and 72°C for 3 minutes. The amplified cDNA library was purified with AMPure beads (Beackman Coulter 10136224) following the manufacturer's recommendations. Amplified cDNA was stored at -20°C for up to a week before use as the IVT template.

RNA Synthesis

In vitro transcription was performed with the NEB HiScribe® T7 High Yield RNA Synthesis Kit (NEB#E2040) in 20 µl of total volume, with 1ug of the previously prepared cDNA product mixed with final concentration 10 mM of: ATP, GTP, UTP, CTP, RNase Inhibitor (Thermofisher 10777019); and 2 µl of the T7 polymerase mix. The reaction was incubated at 42°C for 2 hours in a thermocycler.

IVT RNA purification using phenol chloroform

The resulting IVT was purified by adding equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) (Thermofisher 15593031) to the sample. It was Centrifuged for 5 minutes at 12,000 g at 4°C. The aqueous phase was extracted and transferred to a fresh tube. Equal volume of Chloroform was added to the fresh tube and was centrifuged for 5 minutes at 12,000 g at 4°C (this step was performed twice). The aqueous phase was transferred into a fresh tube with 2.5 volumes 100% ethanol and 0.1 volume 3M sodium acetate. Sample was incubated at -80°C for 1 hour, after that time it was centrifuged for 30 minutes at 12,000 g at 4°C to pellet the RNA. The supernatant was discarded and 500 ul of fresh 70% ethanol was added to the pellet. The sample was centrifuged for 5 minutes at 12,000 g at 4°C and the supernatant was discarded again (this step was repeated twice). The RNA was resuspended in 15ul of H20.

Extended IVT RNA 3′ ends with poly(A)

A maximum of 10 ug of RNA were used for *in vitro* tailing. In 20 µl, the reagents were added in order 1-10 µg IVT RNA, Reaction Buffer to final concentration 1X, 1 mM ATP, 5 units of PolyA polymerase (*E. coli* or yeast as appropriate) (NEB#M0276S). The reaction was incubated at 37°C for either 7 or 30 minutes and stopped by purification with a MEGAclear column (Termofisher AM1908) following the manufacturer's protocol. The purified RNA was optionally enriched for PolyA tails using a MultiMACS mRNA Isolation Kit (Miltenyi Biotec 130-075-201) following the manufacturer’s protocol.

Direct RNA Nanopore Sequencing

RNA was adapted for sequencing using the standard SQK-RNA002 library protocol from ONT and sequenced on R9.4.1 MinION or PromethION flowcells. The resulting fast5 files were basecalled after sequencing with guppy v6.4.6 using the high accuracy model detailed

in the configuration file for MinION (rna_r9.4.1_70bps_hac.cfg) or PromethION (rna_r9.4.1_70bps_hac_prom.cfg) flowcells, as appropriate.

RNA modification detection

The fastq files for the IVT samples were concatenated into a single fastq file, and the directories containing the fast5 files were moved into a single directory to make indexing all the experiments together easier. This was also done for the native samples, but kept in a separate directory from the IVT samples. The two samples were processed as described previously (Leger et al. 2021). Briefly, the fastq files were aligned to the gencode reference transcriptome (v36) (Frankish et al. 2023) concatenated with the SIRVs set 3 reference transcriptome (Lexogen 051.01) using minmap2 (2.20-r1061) (Li 2018) and were filtered and sorted for primary forward alignments using samtools (1.12) (Li et al. 2009). The fast5 files were indexed and the ionic current events in the fast5 files were aligned to the transcriptome reference using nanopolish (0.13.3) (Simpson et al. 2017). The resulting output was collapsed and the position by position pairwise modification detection was done using nanocompore (v1.0.4) (Leger et al. 2021) using context 0. To control for a single RNA modification being called in several neighboring sites, the Nanocompore log10 transformed GMM pvalues for each transcript were peakcalled using scipy.signal.find_peaks with a floor of 2 (corresponding to a pvalue <= 0.01) and a distance of 5 (the kmer width). The resulting peakcalled sites were considered significant for a modification if the absolute value of the log odds ratio was also greater than 0.5 (Leger et al. 2021).

Single Molecule Analysis

Normally, the cluster probabilities for each read is not linked to other positions in the transcript. To map the cluster probabilities at each position to the read of origin, the ionic current intensity and dwell times in the Nanocompore output db were matched to the ionic current intensities and dwell times in the eventalign_collapse.tsv file. In this way the read IDs could be linked to the GMM cluster probabilities. Then, the cluster with more Native sample

label reads as the most probable cluster was selected as the “modified” cluster. A read was considered modified at any position if the modified cluster probability was greater than 0.75. The modification pattern was determined for all transcripts with two or more significant sites by creating a matrix with the number of significant site columns and number of read rows. The matrix values were 0 when the site for that read was unmodified and 1 when modified. The observed counts were determined by counting all the instances of each unique pattern in the matrix. The expected counts were determined by calculating the linear combination of the observed frequencies at each site. The chi2 test of independence was computed using the observed and expected counts for each pattern of modification for each transcript assuming a uniform distribution. The resulting pvalues were multiple testing corrected using the Benjamini-Hochberg FDR procedure.

Additional bioinformatics analysis

Unless otherwise noted, default conditions were used for all analysis tools. Transcriptomic coordinates were converted to genomic coordinates using bedparse (v0.2.3) (Leonardi 2019). Genomic coordinates were translated to transcriptomic coordinates using R library GenomicFeatures function, mapToTranscripts (version 1.46.1). The EMBOSS function, fuzznuc (6.6.0.0) (Rice et al. 2000), was used to find the coordinates of possible METTL3, METTL16, and NSUN6 sequence recognition motifs in either genomic or transcriptomic coordinates. A Nanocompore significant site was inferred as m6A or m5C if the nanocompore 5-mer was within 2 nucleotides of a writer enzyme sequence recognition motif or an RMBase3.0 annotation (Xuan et al. 2024) using bedtools closest (v2.29.1) (Quinlan and Hall 2010). Bedtools closest was also used to determine distances between inferred m6A and m5C sites. Metagene plots were made using the R library, Guitar (2.10.0) (Cui et al. 2016). Nanocompore kmer enrichment analysis was done with Sylamar (version 12-342) (Van Dongen et al. 2008).

DATA ACCESS

All data used in this study can be found at the European Nucleotide Archive using accession number PRJEB81473. Scripts used to analyze the data can be found at https://github.com/nicassiolab/K562_IVT.

COMPETING INTEREST STATEMENT

L.M. and T.L have received financial support from ONT for travel and accommodations to attend and present at ONT events. E.B. is a paid consultant and shareholder of ONT. All other authors declares no conflicts of interest.

ACKNOWLEDGEMENTS

We would like to thank Diego Vozzi and the Genomics Facility at the Center for Human Technologies (CHT@Erzelli) of the Istituto Italiano di Tecnologia (IIT) and Clelia Peano the National Genomics Facility, Human Technopole for conducting the PromethION sequencing for this project. We would like to thank Luca Rotta at the European Institute of Oncology (IEO) Genomic Unit for conducting the MinION sequencing for this project. We would like to thank Salvatore Bianchi for managing all laboratory materials used in this study. We also would like to thank Saul Pierotti for thoughtful discussions about appropriate statistical testing. Figure 1A was created with BioRender.com and released under a Creative Commons Attribution-NonCommercial-NoDerivs (CC-BY-NC-ND) 4.0 International license. This work was supported by grants from the Italian Association for Cancer Research (AIRC) - project IG 2019 (ID. 22851) to F.N. and IG 2020 (ID. 24784) to M.P, from “National Center for Gene Therapy and Drugs based on RNA Technology” (CN00000041) supported by European Union – NextGenerationEU PNRR MUR - M4C2 to F.N., and an AIRC fellowship (ID. 25399) to L.C.T. L.M. was supported by an EMBL ETPOD fellowship.

AUTHOR CONTRIBUTIONS:

L.M and L.C.T. designed the study. L.C.T. and C.R. performed the experiments, L.M., S.M.,

M.F., T.L., and P.M. performed the analysis. L.C.T. and L.M. wrote the initial manuscript.

E.B., F.N., M.P., T.L., M.J.M., T.F. supervised the project and edited the manuscript.

References:

Acera Mateos P, J Sethi A, Ravindran A, Srivastava A, Woodward K, Mahmud S, Kanchi M, Guarnacci M, Xu J, W S Yuen Z, et al. 2024. Prediction of m6A and m5C at single-molecule resolution reveals a transcriptome-wide co-occurrence of RNA modifications. Nat Commun 15: 3899.

Anreiter I, Mir Q, Simpson JT, Janga SC, Soller M. 2021. New Twists in Detecting mRNA Modification Dynamics. Trends Biotechnol 39: 72–89.

Baylin SB, Jones PA. 2011. A decade of exploring the cancer epigenome — biological and translational implications. Nat Rev Cancer 11: 726–734.

Acera Mateos P, J Sethi A, Ravindran A, Srivastava A, Woodward K, Mahmud S, Kanchi M, Guarnacci M, Xu J, W S Yuen Z, et al. 2024. Prediction of m6A and m5C at single-molecule resolution reveals a transcriptome-wide co-occurrence of RNA modifications. *Nat Commun* **15**: 3899.

Anreiter I, Mir Q, Simpson JT, Janga SC, Soller M. 2021. New Twists in Detecting mRNA Modification Dynamics. *Trends Biotechnol* **39**: 72–89.

Baylin SB, Jones PA. 2011. A decade of exploring the cancer epigenome — biological and translational implications. *Nat Rev Cancer* **11**: 726–734.

Begik O, Lucas MC, Pryszcz LP, Ramirez JM, Medina R, Milenkovic I, Cruciani S, Liu H, Vieira HGS, Sas-Chen A, et al. 2021. Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nat Biotechnol* **39**: 1278–1291.

Boccaletto P, Machnicka MA, Purta E, Piątkowski P, Bagiński B, Wirecki TK, de Crécy-Lagard V, Ross R, Limbach PA, Kotter A, et al. 2018. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res* **46**: D303–D307.

Carlile TM, Rojas-Duran MF, Zinshteyn B, Shin H, Bartoli KM, Gilbert WV. 2014. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**: 143–146.

Chandrasekhar C, Kumar PS, Sarma PVGK. 2019. Novel mutations in the kinase domain of BCR-ABL gene causing imatinib resistance in chronic myeloid leukemia patients. *Sci Rep* **9**: 2412.

Cui X, Wei Z, Zhang L, Liu H, Sun L, Zhang S-W, Huang Y, Meng J. 2016. Guitar: An R/Bioconductor Package for Gene Annotation Guided Transcriptomic Analysis of RNA-Related Genomic Features. *BioMed Res Int* **2016**: 1–8.

Delaunay S, Helm M, Frye M. 2024. RNA modifications in physiology and disease: towards clinical applications. *Nat Rev Genet* **25**: 104–122.

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

Esteller M. 2011. Non-coding RNAs in human disease. *Nat Rev Genet* **12**: 861–874.

Frankish A, Carbonell-Sala S, Diekhans M, Jungreis I, Loveland JE, Mudge JM, Sisu C, Wright JC, Arnan C, Barnes I, et al. 2023. GENCODE: reference annotation for the human and mouse genomes in 2023. *Nucleic Acids Res* **51**: D942–D949.

Fu X-D, Ares M. 2014. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat Rev Genet* **15**: 689–701.

Furlan M, Delgado-Tejedor A, Mulroney L, Pelizzola M, Novoa EM, Leonardi T. 2021. Computational methods for RNA modification detection from nanopore direct RNA sequencing data. *RNA Biol* **18**: 31–40.

Garalde DR, Snell EA, Jachimowicz D, Sipos B, Lloyd JH, Bruce M, Pantic N, Admassu T, James P, Warland A, et al. 2018. Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* **15**: 201–206.
Gebauer F, Hentze MW. 2004. Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol* **5**: 827–835.
Gilbert WV, Bell TA, Schaening C. 2016. Messenger RNA modifications: Form, distribution, and function. *Science* **352**: 1408–1412.
Gleeson J, Madugalle SU, McLean C, Bredy TW, De Paoli-Iseppi R, Clark MB. 2024. Isoform-level profiling of m6A epitranscriptomic signatures in human brain. http://biorxiv.org/lookup/doi/10.1101/2024.01.31.578088 (Accessed October 4, 2024).
Harcourt EM, Kietrys AM, Kool ET. 2017. Chemical and structural effects of base modifications in messenger RNA. *Nature* **541**: 339–346.
Hartstock K, Kueck NA, Spacek P, Ovcharenko A, Hüwel S, Cornelissen NV, Bollu A, Dieterich C, Rentmeister A. 2023. MePMe-seq: antibody-free simultaneous m6A and m5C mapping in mRNA by metabolic propargyl labeling and sequencing. *Nat Commun* **14**: 7154.
Henninger JE, Oksuz O, Shrinivas K, Sagi I, LeRoy G, Zheng MM, Andrews JO, Zamudio AV, Lazaris C, Hannett NM, et al. 2021. RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell* **184**: 207-225.e24.
Jady BE. 2001. A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *EMBO J* **20**: 541–551.
Kierzek E, Malgowska M, Lisowiec J, Turner DH, Gdaniec Z, Kierzek R. 2014. The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res* **42**: 3492–3501.
Kirpekar F, Douthwaite S, Roepstorff P. 2000. Mapping posttranscriptional modifications in 5S ribosomal RNA by MALDI mass spectrometry. *RNA* **6**: 296–306.
Leger A, Amaral PP, Pandolfini L, Capitanchik C, Capraro F, Miano V, Migliori V, Toolan-Kerr P, Sideri T, Enright AJ, et al. 2021. RNA modifications detection by comparative Nanopore direct RNA sequencing. *Nat Commun* **12**: 7198.
Leonardi T. 2019. Bedparse: feature extraction from BED files. *J Open Source Softw* **4**: 1228.
Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences ed. I. Birol. *Bioinformatics* **34**: 3094–3100.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinforma Oxf Engl* **25**: 2078–2079.
Li Q, Li X, Tang H, Jiang B, Dou Y, Gorospe M, Wang W. 2017. NSUN2-Mediated m5C Methylation and METTL3/METTL14-Mediated m6A Methylation Cooperatively Enhance p21 Translation. *J Cell Biochem* **118**: 2587–2598.
Linder B, Grozhik AV, Olarerin-George AO, Meydan C, Mason CE, Jaffrey SR. 2015. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods* **12**: 767–772.
Lu L, Zhang X, Zhou Y, Shi Z, Xie X, Zhang X, Gao L, Fu A, Liu C, He B, et al. 2024. Base-resolution m5C profiling across the mammalian transcriptome by bisulfite-free enzyme-assisted chemical labeling approach. *Mol Cell* **84**: 2984-3000.e8.
Maestri S, Furlan M, Mulroney L, Coscujuela Tarrero L, Ugolini C, Dalla Pozza F, Leonardi T, Birney E, Nicassio F, Pelizzola M. 2024. Benchmarking of computational methods for m6A profiling with Nanopore direct RNA sequencing. *Brief Bioinform* **25**: bbae001.
Manrao EA, Derrington IM, Laszlo AH, Langford KW, Hopper MK, Gillgren N, Pavlenok M, Niederweis M, Gundlach JH. 2012. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat Biotechnol* **30**: 349–353.
McCormick CA, Akeson S, Tavakoli S, Bloch D, Klink IN, Jain M, Rouhanifard SH. 2024. Multicellular, IVT-derived, unmodified human transcriptome for nanopore-direct RNA analysis. *GigaByte Hong Kong China* **2024**: gigabyte129.
Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. 2012. Comprehensive

analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**: 1635–1646.
Mikutis S, Gu M, Sendinc E, Hazemi ME, Kiely-Collins H, Aspris D, Vassiliou GS, Shi Y, Tzelepis K, Bernardes GJL. 2020. meCLICK-Seq, a Substrate-Hijacking and RNA Degradation Strategy for the Study of RNA Methylation. *ACS Cent Sci* **6**: 2196–2208.
Monaco PL, Marcel V, Diaz J-J, Catez F. 2018. 2′-O-Methylation of Ribosomal RNA: Towards an Epitranscriptomic Control of Translation? *Biomolecules* **8**: 106.
Morais P, Adachi H, Yu Y-T. 2021. The Critical Contribution of Pseudouridine to mRNA COVID-19 Vaccines. *Front Cell Dev Biol* **9**: 789427.
Motorin Y, Helm M. 2010. tRNA Stabilization by Modified Nucleotides. *Biochemistry* **49**: 4934–4944.
National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Division on Earth and Life Studies; Board on Health Sciences Policy; Board on Life Sciences; Toward Sequencing and Mapping of RNA Modifications Committee. 2024. *Charting a Future for Sequencing RNA and Its Modifications: A New Era for Biology and Medicine*. National Academies Press (US), Washington (DC) http://www.ncbi.nlm.nih.gov/books/NBK606036/ (Accessed October 7, 2024).
Pardi N, Krammer F. 2024. mRNA vaccines for infectious diseases — advances, challenges and opportunities. *Nat Rev Drug Discov*. https://www.nature.com/articles/s41573-024-01042-y (Accessed October 7, 2024).
Pardo-Palacios FJ, Wang D, Reese F, Diekhans M, Carbonell-Sala S, Williams B, Loveland JE, De María M, Adams MS, Balderrama-Gutierrez G, et al. 2024. Systematic assessment of long-read RNA-seq methods for transcript identification and quantification. *Nat Methods* **21**: 1349–1363.
Ping X-L, Sun B-F, Wang L, Xiao W, Yang X, Wang W-J, Adhikari S, Shi Y, Lv Y, Chen Y-S, et al. 2014. Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res* **24**: 177–189.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.
Rice P, Longden I, Bleasby A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet* **16**: 276–277.
Roundtree IA, Evans ME, Pan T, He C. 2017. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **169**: 1187–1200.
Schumann U, Zhang H-N, Sibbritt T, Pan A, Horvath A, Gross S, Clark SJ, Yang L, Preiss T. 2020. Multiple links between 5-methylcytosine content of mRNA and translation. *BMC Biol* **18**: 40.
Selmi T, Hussain S, Dietmann S, Heiß M, Borland K, Flad S, Carter J-M, Dennison R, Huang Y-L, Kellner S, et al. 2021. Sequence- and structure-specific cytosine-5 mRNA methylation by NSUN6. *Nucleic Acids Res* **49**: 1006–1022.
Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. 2017. Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods* **14**: 407–410.
Sloan KE, Warda AS, Sharma S, Entian K-D, Lafontaine DLJ, Bohnsack MT. 2017. Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol* **14**: 1138–1152.
Song L-F, Deng Z-H, Gong Z-Y, Li L-L, Li B-Z. 2021. Large-Scale de novo Oligonucleotide Synthesis for Whole-Genome Synthesis and Data Storage: Challenges and Opportunities. *Front Bioeng Biotechnol* **9**: 689797.
Stephenson W, Razaghi R, Busan S, Weeks KM, Timp W, Smibert P. 2022. Direct detection of RNA modifications and structure using single-molecule nanopore sequencing. *Cell Genomics* **2**: 100097.
Svitkin YV, Gingras A-C, Sonenberg N. 2022. Membrane-dependent relief of translation elongation arrest on pseudouridine- and N1-methyl-pseudouridine-modified mRNAs. *Nucleic Acids Res* **50**: 7202–7215.
Tavakoli S, Nabizadeh M, Makhamreh A, Gamper H, McCormick CA, Rezapour NK, Hou Y-M, Wanunu M, Rouhanifard SH. 2023. Semi-quantitative detection of pseudouridine

modifications and type I/II hypermodifications in human mRNAs using direct long-read sequencing. *Nat Commun* **14**: 334.
Uzonyi A, Dierks D, Nir R, Kwon OS, Toth U, Barbosa I, Burel C, Brandis A, Rossmanith W, Le Hir H, et al. 2023. Exclusion of m6A from splice-site proximal regions by the exon junction complex dictates m6A topologies and mRNA stability. *Mol Cell* **83**: 237-251.e7.
Van Dongen S, Abreu-Goodger C, Enright AJ. 2008. Detecting microRNA binding and siRNA off-target effects from expression data. *Nat Methods* **5**: 1023–1025.
Wan Y, Qu K, Zhang QC, Flynn RA, Manor O, Ouyang Z, Zhang J, Spitale RC, Snyder MP, Segal E, et al. 2014. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* **505**: 706–709.
Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, Fu Y, Parisien M, Dai Q, Jia G, et al. 2014. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**: 117–120.
Wang X, Zhao BS, Roundtree IA, Lu Z, Han D, Ma H, Weng X, Chen K, Shi H, He C. 2015. N6-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **161**: 1388–1399.
Workman RE, Tang AD, Tang PS, Jain M, Tyson JR, Razaghi R, Zuzarte PC, Gilpatrick T, Payne A, Quick J, et al. 2019. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat Methods* **16**: 1297–1305.
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, et al. 2016. Nuclear m 6 A Reader YTHDC1 Regulates mRNA Splicing. *Mol Cell* **61**: 507–519.
Xuan J, Chen L, Chen Z, Pang J, Huang J, Lin J, Zheng L, Li B, Qu L, Yang J. 2024. RMBase v3.0: decode the landscape, mechanisms and functions of RNA modifications. *Nucleic Acids Res* **52**: D273–D284.
Zhang Z, Chen T, Chen H-X, Xie Y-Y, Chen L-Q, Zhao Y-L, Liu B-D, Jin L, Zhang W, Liu C, et al. 2021. Systematic calibration of epitranscriptomic maps using a synthetic modification-free RNA library. *Nat Methods* **18**: 1213–1222.