communications medicine

A Nature Portfolio journal

Article

https://doi.org/10.1038/s43856-025-01235-5

Nanopore based RNA methylation profiling of a circulating lung cancer biomarker

Check for updates

Marta Sanchez-Delgado[1,3], Maurice Frank[1,3], Tomáš Šišmiš[2], Mustafa Kahraman[1], Alberto Daniel-Moreno[1], Emmika Mummery[1], Jessika Ceiler[1], Jasmin Skottke[1], Carla Bieg-Salazar[1], Franziska Hinkfoth[1], Christina Rudolf[1], Ronja Weiblen[1], Kaja Tikk[1], Tobias Sikosek[1], Bruno R. Steinkraus[1], Rastislav Horos[1], Michal Urda[2] & Timothy Rajakumar[1]

Abstract

Background The characterisation of circulating nucleic acid biomarkers in liquid biopsy based diagnostics holds great potential to transform the landscape of early cancer detection and screening. These tests are increasingly incorporating information beyond the primary sequence, to include epigenetic, fragmentomic, and other chemical properties to boost performance. Chemical modifications of RNA offer a rich, and currently underutilised source of biomarker signal. We aimed to develop a single-molecule method to profile 2′-O-methylation of a ribosomal RNA fragment previously linked to lung cancer, and to evaluate its diagnostic value in blood.

Methods We have designed a targeted capture strategy that ligates structure-guided adapters to a ~22-nucleotide ribosomal RNA fragment to enable sequencing of native molecules on a nanopore platform. Raw ionic-current signals were used to train and validate machine-learning classifiers to detect methylation states of synthetic oligonucleotides and cell culture derived ribosomal RNA fragments. Finally, we applied these methods to liquid biopsy samples collected from a 43-patient cohort of individuals undergoing investigation for suspected lung cancer.

Results Here we show that single molecules of the target fragment are sequenced, and their methylation states can be accurately (92%) and quantitatively (Pearson r = 0.997) measured. In clinical liquid biopsy samples, it reveals a differential pattern of methylation in lung cancer that yields a diagnostic classifier with an area under the receiver-operating characteristic curve of 0.84.

Conclusions This approach enables direct, single-molecule methylation profiling of small RNAs in blood and identifies a lung cancer–associated methylation pattern with diagnostic potential. It is readily compatible with multi-modal liquid biopsy assays to enhance performance.

Plain language summary

Blood tests for the early detection of cancer show great promise. These rely on the analysis of millions of molecules in the blood and the detection of differences that indicate the presence of cancer. RNA molecules are typically assessed by reading the letters of their genetic code (A, U, C, G), however we now realise that these letters can undergo chemical modifications, akin to changing their font. We have built a method that can read these molecules and identify their font using artificial intelligence. We used this method to investigate one RNA molecule in blood samples and found a difference between the fonts used in healthy compared to patients with lung cancer. This method could add a new layer to blood-based cancer tests and may help us find lung cancer earlier.

Liquid biopsy is poised to transform cancer screening through the non-invasive detection of cancer biomarkers such as circulating tumor DNA (ctDNA)[1,2], cell-free RNA (cfRNA)[3–5], proteins[6], and auto-antibodies[7,8]. The analysis of DNA-based biomarkers has increasingly focused on properties beyond the primary sequence, including epigenetic modifications[9,10], fragmentation patterns[11], and end motifs[12], to increase the sensitivity and specificity of cancer detection. Similar epitranscriptomic characteristics, including RNA modifications, are beginning to be recognised as potential cancer biomarkers[13–16]. Tools to explore the diverse complexity of RNA modifications, including over one hundred

[1]Hummingbird Diagnostics GmbH, Heidelberg, Germany. [2]Department of Pneumology and Phtiseology, University Hospital and Polyclinic F.D. Roosevelt, Banská Bystrica, Slovakia. [3]These authors contributed equally: Marta Sanchez-Delgado, Maurice Frank. e-mail: publication@hb-dx.com

different modified bases[17], may unlock a rich new source for biomarker discovery.

Direct RNA sequencing (DRS) on the Oxford Nanopore (ONT) platform enables such deep characterisation, by natively analysing single RNA molecules, complete with all modifications intact[18]. This data is captured as electrical current over time as an RNA molecule passes through a nanopore and is translated to biologically relevant information algorithmically. Ever-evolving machine-learning models have brought about improvements in canonical basecalling accuracy from ~86%[19] to ~97%[20]. Similar advances have occurred in methods developed for the detection of RNA modifications. Initially, comparative analyses enabled the identification of site-level modifications using differences in nanopore raw signal data between reference and modification-deficient samples[21–24]. More recently, machine-learning models have been trained for the single-nucleotide, single-molecule resolution detection of RNA modifications, including m6A, m5C, and pseudouridine[25–29]. Such techniques will further our understanding of the extent and role of RNA modifications.

The ribosomal RNA (rRNA) represents a promising source of epitranscriptomic information, given that it is heavily modified at over 228 sites with 14 different modifications[30–32]. rRNA fragments (rRFs) can be generated in a non-random manner with functional importance[33] and can serve as small RNA biomarkers of cancer[34]. Here, we have focused on the characterisation of a specific 22 nt lung cancer rRF ($28S^{4014–4035}$) that can be modified by 2′-O-Methylation (Nm) at two distinct sites (G7 and C19) (Fig. 1a). We have developed methods for the targeted capture of this rRF into a Nanopore sequencing library and the algorithmic classification of its Nm modification state. This constitutes the shortest RNA species successfully sequenced to date on the ONT platform, as well as the first single-molecule characterisation of Nm methylation dynamics. We show that targeted ONT sequencing of $28S^{4014–4035}$ in biological samples reflects methylation-deficient mutant phenotypes and reveal lung cancer-associated differential $28S^{4014–4035}$ Nm methylation patterns in cfRNA from a clinical cohort. Together, we envision that this method will enable the incorporation of small RNA modification measurements into multi-omic biomarker development that can enhance the performance of liquid biopsy tests.

Methods

Preparation of sequencing library adapters

Four different sets of single-stranded RNA and/or DNA oligonucleotides were used (IDT, Coralville, IA):

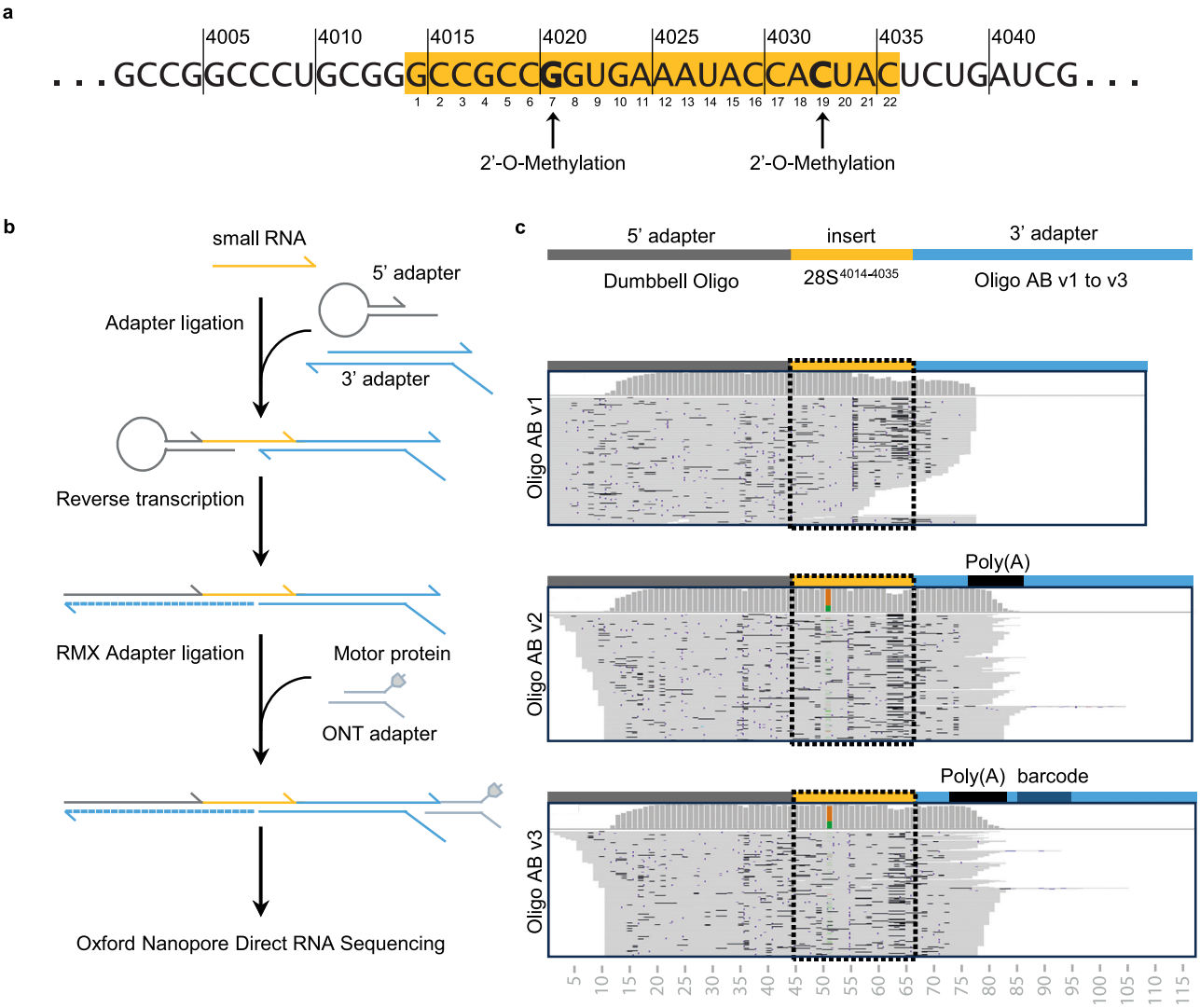

Fig. 1 | A library preparation method for small RNA DRS. a Sequence of the rRF $28S^{4014–4035}$ (highlighted in yellow) within the context of the human 28S rRNA, marking the Nm methylated positions. **b** Schematic of the library preparation method, illustrating the capture of a target small RNA by ligation to specific 5′ and 3′ adapters, reverse transcription, ligation of ONT adapters and motor protein, and finally sequencing. **c** Schematic overview of the DRS library structure, and IGV snapshots demonstrating the mapping of small RNA DRS reads to their respective references, and the impact of evolving adapter design.

(1) Four different RNA oligonucleotides matching the sequence of $28S^{4014–4035}$. All of them contain a 5′-phosphate group, and they differ in their Nm methylation (GC, GmC, GCm, GmCm).
(2) Dumbbell Oligo (5′ adapter) as an RNA oligonucleotide.
(3) Oligo A (part of 3′ adapter) is a single-stranded RNA and DNA hybrid molecule with a 5′-phosphate group.
(4) Oligo B (3′ adapter and RT primer) is a single-stranded DNA oligonucleotide, which was modified with LNAs in v2 and v3.

As described in the ONT DRS protocol, annealing buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl) was used to prepare the 3′ adapter (1.4 μM final working concentration, mixing Oligo A and Oligo B). The annealing buffer was also used to dilute the 5′ adapter to result in a 2.5 μM working concentration. Both adapters were incubated at 82 °C for 2 min followed by a ramp down of 0.1 °C/s to 4 °C and stored at −20 °C.

Direct RNA sequencing library preparation for synthetic oligonucleotides

Four different RNA oligonucleotides matching the sequence of $28S^{4014–4035}$, with different combinations of Nm methylation (GC, GCm, GmC, GmCm) used to train AI classification models, were sequenced on Flongle Flow Cells (R9.4.1). Libraries were prepared using a modified Flongle Direct RNA Library preparation protocol[35].

For the first RNA ligation, 0.025 μM of $28S^{4014–4035}$ synthetic oligo was used for ligation with 0.5 μM of 5′ adapter (IDT), 5% Polyethyleneglycol 8000 (w/v) (Jena Bioscience GmbH), 1x RNA ligation buffer, 10 units of T4 Rnl2, 1 mM ATP, and 0.14 μM of the corresponding barcoded 3′ adapter. The ligation mix was incubated at 37 °C for 60 min with the lid heated to 45 °C, then at 16 °C for 2 h and 12 °C overnight. The previous adapter-ligated RNA was used to continue with the reverse transcription step using SuperScript III reverse transcriptase (Thermo Fisher Scientific, DE) as described in the DRS protocol (adjusting the sample volume with nuclease-free water as indicated in the DRS protocol).

After reverse transcription and heat inactivation, separately barcoded samples were mixed. Using RNAClean XP beads (Beckman Coulter Agencourt) at a beads:cDNA volume ratio of 2:1, cleanup proceeded as described in the ONT instructions for DRS. The beads were eluted in 12.5 μl of nuclease-free water, and the RMX Adapter ligation was prepared in ½ of the recommended volumes indicated in the DRS protocol. The ligation mix was incubated at room temperature for 20 min.

Prior to continuing with the 2nd beads cleanup, we adjusted the ligation volume to 40 μl and used RNAClean XP beads at a beads:cDNA volume ratio of 0.4:1 as described in the ONT DRS protocol. Finally, the library was eluted in 7.5 μl of Elution Buffer (EB) and mixed with 7.5 μl of nuclease-free water and 15 μl of RNA Running Buffer (RRB) to prepare the final library mix.

For loading the library into the Flongle Flow cell (R9.4.1), first we slowly load a mix of 117 μl Flush Buffer with 3 μl Flush Tether and then the previously prepared library mix.

Nanopore sequencing

All samples were sequenced with default MinKNOW parameters for DRS libraries prepared with the Direct RNA Sequencing Kit (SQK-RNA002), ensuring that output bulk sequencing files in fast5 format were saved. Subsequently, the playback feature of MinKNOW was used to replay the bulk fast5 file using a custom short RNA config file (sequencing_MIN106_RNA_short.toml) to more efficiently segment and capture small RNA reads[36]. These reads in fast5 format were used for further downstream analysis and basecalling.

Nanopore data preprocessing

We first basecalled the raw fast5 data using the ONT Guppy basecaller and rna_r9.4.1_70bps_sup model[20], discarding reads with low quality (mean read PHRED < 7). The remaining reads were mapped against a custom reference (5′ adapter + $28S^{4014–4035}$ + 3′ adapter) using the Burrows-Wheeler Aligner[37]. Only reads mapping to the full-length library were retained.

Raw data associated with reads passing the above QC criteria were next split into a 5′ RNA segment and 3′ DNA segment based on the automated identification of the RNA/DNA transition and characteristic current shift using a dynamic programming algorithm for change point detection (Ruptures[38]). Both segments were normalised using the Median Absolute Deviation and scaled to a fixed length (10,000) through linear interpolation. This is required as the raw current data varies in temporal granularity and may not capture the exact beginning and end of the strand, so interpolation approximates the data to a similar scale and enables a consistent input size for deep learning models. Finally, we removed amplitude spikes from the raw data by eliminating values that fall outside three standard deviations from the median, replacing them with interpolated values from a linear rolling window.

Nanopore data machine learning

We trained two ResNet-based 1D deep learning models on the normalised, scaled raw datasets:

1. Predicting the modification status of the target RNA insert ($28S^{4014–4035}$, in either GC, GmC, GCm, or GmCm form).
2. Predicting the barcode within the DNA portion of the 3′ adapter (barcode 1–6).

Preprocessing of the raw data was used to produce a dataset of equal-length 1D signals (10,000 features per read) that was used for training the two deep learning models. For the two tasks of predicting methylation status and barcoding, we collected datasets of 190,891 and 255,759 reads, respectively (Supplementary Tables 3 and 4). Both datasets were randomly split into training (80%) and testing (20%) sets.

The two models have the same architecture, excluding the output layer, but do not share any weights. The model architectures consist of eight residual blocks, each comprising three 1D convolutional layers, followed by batch normalisation and ReLU activation. As residual blocks, their outputs are added back to the input of the block. The channel width of the convolutions increases from 20 in the first block to 67 in the last block, across three steps. Before the first block, there is an additional 1D convolutional layer with 20 channels, followed by batch normalisation, ReLU activation, and a pooling layer that reduces the temporal size of the input. After the last block, the remaining temporal dimension is averaged to produce an embedding vector of size 67. A linear classifier uses these embeddings to output logits for each class of the specific training task. Ignoring the size of the final classifier, both models have 66785 parameters. This architecture is based on the SquiggleNet architecture[39]. The cross-entropy loss of the models was optimised using the Adam optimiser with a learning rate of 1e-2 and a batch size of 1024. Training was done until loss convergence.

Cell lines and culture

HCC-827 human lung cancer cell line was obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ; No.: ACC 566) and cultured at 37 °C with 5% CO_2 in RPMI (Thermo Fisher Scientific) supplemented with 20% FBS (Thermo Fisher Scientific) and 1% Penicillin-Streptomycin (Thermo Fisher Scientific).

Knockout generation

CRISPR/Cas9 was used to generate SNORD102 and SNORD75 single-cell-colony knockouts from HCC-827 cells. Chemically modified sgRNAs (IDT; Supplementary Table 7) and Cas9 ribonucleoprotein (RNP; IDT) were incubated at a molar ratio of 2:1 (90–45 pmol) at room temperature for 15 min. After complex formation, 80,000 HCC-827 cells were transfected using the 10 μl Neon Transfection Kit (Thermo Fisher Scientific) with the following electroporation settings: 1400 V, 20 ms, 2 pulses[40].

Indel analysis

Genomic DNA was isolated 5 days post-transfection using the NucleoSpin Tissue Kit according to the manufacturer's instructions (MACHEREY-NAGEL). The target regions were amplified using GoTaq Green Master Mix (Promega). Primers for each target region are listed in Supplementary Table 8. PCR products were purified using the DNA Clean & Concentrator (Zymo Research). Indel rates for each target were evaluated using the web tool "ICE" (Inference of CRISPR Edits) after Sanger sequencing of the purified PCR products. Next, single-cell sorting was performed, and the most suitable colonies were selected based on their indel percentage and morphology.

RNA extraction from cell medium

A total of 1 million HCC-827 cells were seeded in T75 flasks, and RNA was extracted from 10 ml of cell medium (which had been incubated with growing cells for at least 24 h), using the Quick-cfRNA Serum and Plasma Kit (Zymo Research) according to the manufacturer's protocol with a few modifications. Initial centrifugation was performed in 15 ml conical tubes, at 4347 × *g* for 15 min to remove any cell debris and precipitate. The samples were eluted in 15 µl of DNase/RNase-Free Water.

Direct RNA sequencing library preparation for biological samples

The ONT DRS protocol (SQK-RNA002) was used with modifications as described below.

Extracted RNA was first PNK-treated to ensure 5′ phosphorylation. A maximum of 500 ng of total RNA (previously incubated at 70 °C for 2 min) was mixed with 10 units of T4 PNK (NEB), T4 PNK Reaction Buffer (1X), and 1 mM of ATP. The samples were incubated at 37 °C for 30 min, and heat-inactivated at 65 °C for 20 min. Since cfRNA samples have low concentration (less than 10 ng/µl), we used the maximum volume of RNA possible for the PNK conversion, in our case, 5.65 µl.

The first ligation was performed with PNK-treated RNA and 0.5 µM of 5′ adapter (IDT), 5% Polyethyleneglycol 8000 (w/v) (Jena Bioscience GmbH), 1x RNA ligation buffer, 10 units of T4 Rnl2, 1 mM ATP, and 0.14 µM of the corresponding barcoded 3′ adapter. The ligation mix was incubated at 37 °C for 60 min with the lid heated to 45 °C, then at 16 °C for 2 h and 12 °C overnight. The adapter-ligated RNA was used to continue to the reverse transcription step using SuperScript III reverse transcriptase (Thermo Fisher Scientific, DE) as described in the ONT DRS protocol (adjusting the sample volume with Nuclease-free water) and followed by heat inactivation of T4 Rnl2 at 80 °C for 5 min.

After reverse transcription and heat inactivation, separately barcoded samples were mixed for multiplexing. Cleanup was performed as per the DRS protocol using RNAClean XP beads (Beckman Coulter Agencourt) at a beads:cDNA volume ratio of 2:1. If the concentration of the samples was too low to quantify, bead elution was performed with 23 µl of nuclease-free water without adding more water in the proceeding RMX adapter ligation described in the DRS protocol.

Both the cfRNA from the study participants and the medium from HCC-827 were sequenced in MinION Flow Cells (R9.4.1) from Oxford Nanopore.

Clinical study participants

The study was approved by and performed at the University Hospital with Polyclinic F.D.Roosevelt in Banská Bystrica, Slovakia (ethical approval number 26/2023 from August 30, 2023). The study was registered at the German Clinical Trials Register (DRKS) under the number DRKS00032631 on Sept 8, 2023. The recruited participants were 55–80 years old, current or ex-smokers with at least 20 pack-years smoking history, and with nicotine absence for less than 15 years (if ex-smoker). All participants provided written informed consent.

Clinical sample collection, processing, and extraction

Ten milliliters of blood was collected in Streck RNA Complete BCT tubes and processed by centrifugation within 4 h according to the manufacturer's instructions. Plasma was aliquoted and stored at −80 °C until extraction. RNA was extracted from 0.5 mL plasma using the Quick-cfRNA Serum and Plasma Kit (Zymo Research) according to the manufacturer's instructions, eluted in 18 µl of RNase-free water, and stored at −80 °C.

Statistics and reproducibility

All statistical analyses were performed in Python. The specific sample sizes, statistical tests, and findings are reported in the respective figure legends.

Ethics approval and consent to participate

All research has been performed in accordance with the Declaration of Helsinki. Ethical permits are in place for all studies, and all participants consented to partake.

Results

Small RNA sequencing on the Oxford Nanopore platform

DRS on the ONT platform has primarily been applied to longer RNA molecules[41]. The sequencing of shorter molecules has proved challenging; however, recently, methods have been developed to enable DRS of shorter non-coding RNAs, including snoRNAs and snRNAs[42], and ~75–100 nucleotide long tRNAs[36,43]. tRNA sequencing methods are enabled by the ligation of 5′ and 3′ adapter sequences to embed the small RNA of interest within a longer molecule amenable to sequencing.

We have extended the capability of ONT DRS to the smallest RNA species (~22 nt rRFs) with a structured adapter ligation method that captures target small RNAs into an ONT DRS-compatible library (Fig. 1b). This approach utilises a 5′ adapter that adopts a stem-loop structure based on the Dumbbell (DB) PCR method[44] to enforce ligation to target small RNAs with single nucleotide discrimination at the 5′ end, and a target specific double stranded Nanopore Reverse Transcription Adapter (Oligo AB) for ligation to the 3′ end. All work has been conducted using the ONT RNA002 chemistry and R9.4.1 flow cells.

Initial method development and optimisations were conducted by designing adapters to ligate to a synthetic RNA oligo matching the sequence of the $28S^{4014–4035}$ rRF (Fig. 1a). The first 3′ adapter (Oligo AB v1) iteration was designed as per ONT recommendation to simply replace the 10 nt 3′ overhang with a target specific reverse complementary sequence (Supplementary Fig. 1a). The 5′ adapter (Dumbbell Oligo) was designed to form a stem-loop structure with an 8 nt 5′ overhang (Supplementary Fig. 1). Qualitatively efficient single-step ligation of both Dumbbell Oligo and Oligo AB v1 adapters occurred (Supplementary Fig. 2) to create a substrate compatible for subsequent ligation of the Nanopore adapter and motor protein complex (RMX) and DRS. Following DRS, basecalling, and alignment, we observed approximately 40% of reads to pass QC with a PHRED > 7 and alignment to the expected library sequence (Supplementary Table 1). QC passed reads displayed deep coverage of the 5′ adapter indicative of efficient ligation; however, incomplete coverage of the target 3′ and adapter regions (Fig. 1c, upper panel). We reasoned that the basecalling algorithms had likely been predominantly trained on poly(A) containing mRNAs, and the absence of this structure in the Oligo AB v1 3′ adapter design may have hindered the accurate identification of the 3′ end and basecalling of this region. To address this, we designed a second iteration of the 3′ adapter (Oligo AB v2) containing a 10 nt poly(A) tract within the RNA portion of the adapter (Supplementary Fig. 1b). Additionally, the second generation of 3′ adapter contained LNA modifications at 3 positions within the 3′ overhang to increase affinity for the target small RNA (Supplementary Fig. 1b). While the proportion of reads passing QC remained approximately stable (Supplementary Table 1), there was a marked improvement in the depth of coverage of the target 3′ region (Fig. 1c, middle panel). A third and final 3′ adapter design (Oligo AB v3) added a 10 nt DNA barcode 3′ of the poly(A) tract, within the DNA portion of the adapter, in a similar manner to the barcoding method described by Smith et al. (Supplementary Fig. 1c)[45]. DRS sequencing of a library created with Oligo AB v3 adapters revealed minimal impact of the additional barcode with stable QC metrics (Supplementary Table 1) and maintained accurate sequencing and

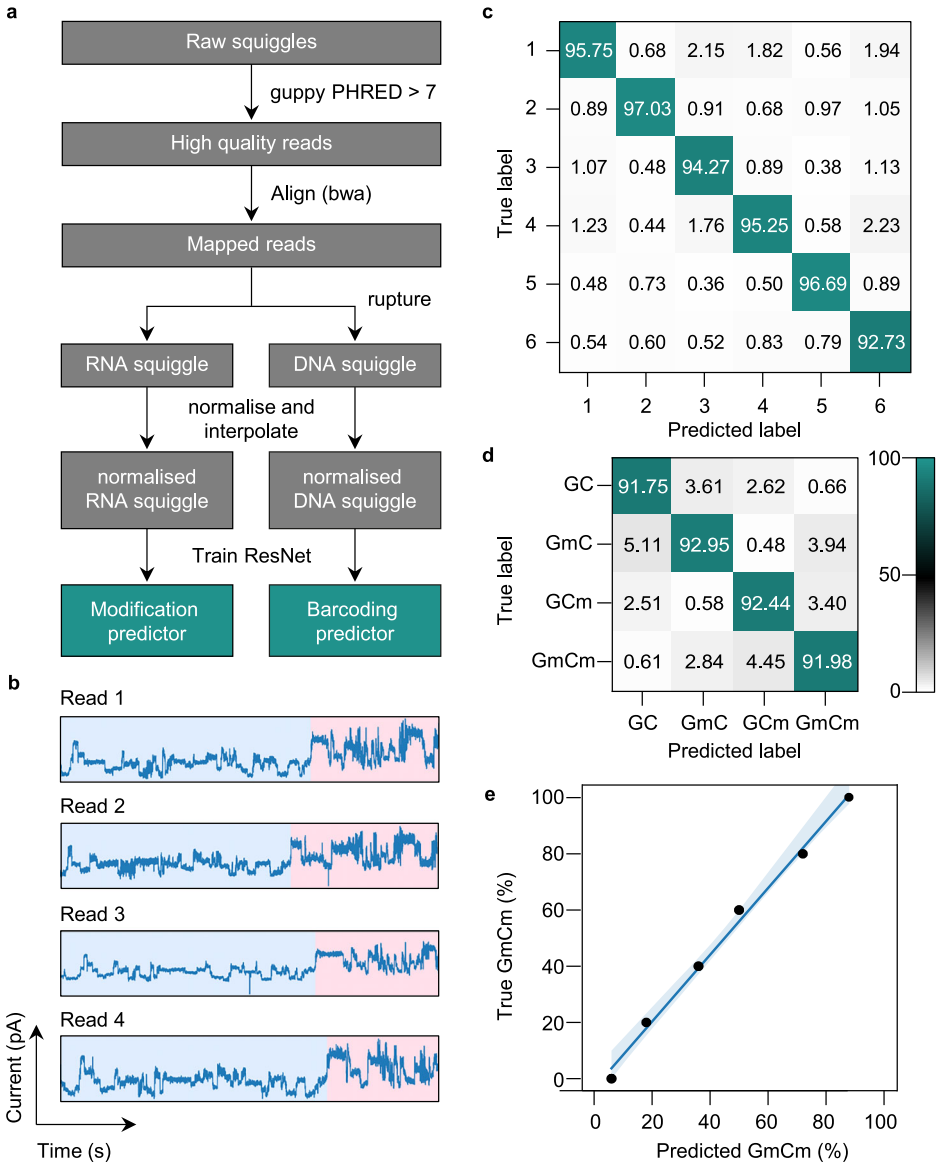

Fig. 2 | Training AI models for RNA methylation and barcode classification. a Schematic flowchart detailing the steps for processing raw DRS data and training of RNA modification and barcoding models. **b** Representative raw data for four randomly chosen reads demonstrates the automated segmentation of reads into a 5′ RNA and 3′ DNA portion. **c** Confusion matrix for the prediction of barcode labels (1–6) in held-out test data. **d** Confusion matrix for the prediction of 28S$^{4014–4035}$ Nm methylation status in held-out test data. **e** Standard curve for GmCm proportion, diluted in a background of GCm. The solid line is the result of linear regression with shading to indicate the 95% confidence interval.

depth of insert coverage (Fig. 1c, lower panel). The emergence of an SNV at position 7 of the 28S$^{4014–4035}$ insert is observed with Oligo AB v2 and v3 adapters. Given that the insert sequence remains the same, this difference may be explained by changes in the nanopore current signal and error profile from the altered surrounding adapter context.

Together, our protocol with single-step ligation of both a Dumbbell Oligo 5′ adapter and Oligo AB v3 3′ adapter efficiently captured a synthetic 22 nt RNA oligo to create a library compatible with DRS with the potential to capture information regarding RNA modifications.

AI classification models for small RNA methylation status and barcode demultiplexing

Following the proof of principle that ~22 nt small RNAs could be sequenced with ONT DRS, we sought to extend the utility of this by training AI models to enable the methylation status classification of a reported small RNA cancer biomarker, and the classification of 3′ adapter barcodes for multiplexed DRS. Given the intended use of this method for sequencing of small RNA targets with known patterns of modification, we have approached the AI task as one of multiclass classification of entire DRS reads as a whole, rather than single-nucleotide resolution modified basecalling. We intentionally designed these as two independent and modular classification tasks to enable the interchangeability of different target small RNA libraries with a fixed set of barcodes that could be demultiplexed without the need to retrain the barcode models from scratch.

AI model development was performed using synthetic rRF 28S$^{4014–4035}$ and a set of six 3′-adapter barcode sequences (Supplementary Table 2). Given that the 22 nt rRF 28S$^{4014–4035}$ contains two reported Nm modification sites at positions G7 and C19, it can exist in one of four distinct forms denoted GC (unmethylated), GmC (G7 methylated), GCm (C19 methylated), and GmCm (double methylated). A total of 24 DRS libraries were created using the above-described method to create a ground truth dataset containing synthetic 28S$^{4014–4035}$ in each of its possible four states of Nm methylation, each with six different 3′-adapter barcodes. These were sequenced on individual ONT Flongle flow cells (R9.4.1) to generate the raw data to train the AI classification models (Fig. 2a).

Initial QC was performed on the raw nanopore data for each library, consisting of basecalling to select reads of acceptable quality (mean PHRED > 7) and alignment to select for reads of the expected library structure containing 5′ adapter, target small RNA insert, and 3′ adapter (Supplementary Tables 3 and 4). Passed reads were then segmented into a 5′ RNA portion containing the small RNA insert and a 3′ DNA portion containing the barcode, by exploiting the characteristic shift in nanopore raw current intensity between RNA and DNA (Fig. 2b), using a dynamic programming point detection approach[38]. This segmentation enabled the

subsequent independent training of AI models for both small RNA insert modification status and barcode classification.

The read level, raw current data, corresponding to the small RNA and barcode containing segments, were then further processed to scale to a fixed length of 10,000 and normalised. This processed data was used to train separate 1D ResNet models[39], for the classification of small RNA modification ($28S^{4014–4035}$ methylation) and barcode status (further details in the methods) (Supplementary Fig. 3). An 80:20 data split was used to train and test performance. We observe a test set accuracy of 92% and 95% for the small RNA methylation and barcode status models, respectively (Fig. 2c, d). In order to control for any data leak from the flowcell batch effect and to explore the quantitative performance of this classifier, we created a multiplexed standard curve of synthetic rRF $28S^{4014–4035}$ (GmCm in GCm) and sequenced on a single flowcell. We observe a strong and significant correlation between the predicted and ground truth GmCm methylation fraction (Pearson $r = 0.9971$, $p = 1.24e\text{-}05$) (Fig. 2e).

This confirms the potential of DRS as a means to profile small RNA modification status in a quantitative and multiplexed manner on the ONT platform.

Detection of small RNA modifications in biological samples

To further explore the utility of this method, we sought to test its ability to quantify small RNA modification status in biological samples. We used an HCC-827 cell culture model, which has been previously shown to release the rRF $28S^{4014–4035}$ at high levels into its conditioned growth medium[34], to create sequencing libraries for the targeted DRS of this small RNA fragment.

The efficiency of library preparation from RNA extracted from HCC-827 cell culture supernatant was greatly reduced relative to earlier libraries prepared with synthetic RNA, as reflected by the reduced proportion of approximately 0.4% of reads passing QC (Supplementary Table 5). Despite the high relative expression of rRF $28S^{4014–4035}$ in conditioned media, the absolute concentration of RNA obtained from extractions is very low (<10 ng/µl), posing a challenge for its efficient capture into sequencing libraries. To negate this low efficiency, we have processed all biological samples on the larger throughput Minion flow cells (R9.4.1).

It has been previously reported that the Nm modification frequency of positions G7 and C19 within this rRNA region are 83% and 100% respectively[32]. These Nm modifications are deposited in a site-specific manner under the guidance of complementary C/D box snoRNAs (SNORDs) (Fig. 3a). In the conditioned media of wild-type HCC-827 cells, we observe a similar tendency in the frequency of Nm modification at position G7 of 60.4% (GmC and GmCm) and C19 of 84.6% (GCm and GmCm) (Fig. 3b). To perturb this, we used CRISPR/Cas9 to create mutant HCC-827 clones with disrupted guide sequences of either SNORD102 or SNORD75 (Supplementary Fig. 4), which are the SNORDs responsible for the 2′-O-methylation of G7 and C19 respectively. Targeted Nanopore sequencing of the $28S^{4014–4035}$ rRF from the conditioned media of these mutant cell lines reveals the expected reduction of Nm at positions G7 and C19 in SNORD102 and SNORD75 mutants, respectively (Fig. 3b).

Detection of small RNA modifications in clinical samples and use as a cancer diagnostic

Finally, we applied the method developed here to quantify the Nm methylation status of the $28S^{4014–4035}$ rRF in a clinical lung cancer cohort. Forty-three patients aged 55–80, with a smoking history of at least 20 pack-years, and who are either current or ex-smokers having quit within the last 15 years, were prospectively recruited from new referrals with a suspicion of lung cancer (Table 1). Before any invasive procedures or definitive diagnosis, participants donated 10 ml blood in Streck RNA tubes, from which 0.5 ml of plasma was used for RNA extraction and library preparation for targeted DRS according to the small RNA sequencing method described earlier.

The circulating levels of $28S^{4014–4035}$ are higher in the blood of lung cancer patients compared to control patients[34]; however, the Nm methylation status of this fragment has not been previously explored. We demonstrate the successful capture and Nanopore sequencing of this rRF from plasma samples, albeit at low efficiencies (Supplementary Table 5) and a trend towards higher total read counts from cancer compared to control samples (Supplementary Fig. 5). We observed a significant increase in the proportion of $28S^{4014–4035}$ with a GmCm methylation pattern ($p = 1.05e\text{-}03$, Cohen's d effect size = 1.33) in the plasma of lung cancer patients compared to controls (Fig. 3c). This is accompanied by a corresponding significant decrease in $28S^{4014–4035}$ with GC, GmC and GCm methylation patterns (Fig. 3c). The shifting pattern of $28S^{4014–4035}$ methylation is observed even in samples from patients with early stage (I and II) lung cancer (Supplementary Fig. 6a), however these are of limited number in the cohort presented here. Furthermore, these changes are consistent between NSCLC and SCLC lung cancer (Supplementary Fig. 6b) and between histological subtypes of NSCLC (Supplementary Fig. 6c).

We further assessed the diagnostic potential of $28S^{4014–4035}$ GmCm methylation in a logistic regression model that included clinical covariates (age, sex, and smoking history in pack-years). Both $28S^{4014–4035}$ GmCm methylation and smoking history were independently and significantly associated with lung cancer diagnosis (Supplementary Table 6), suggesting that the methylation status has diagnostic value beyond known risk factors.

Taken together, $28S^{4014–4035}$ GmCm methylation proportion serves as a blood-based biomarker for lung cancer that can distinguish cancer from control patients with a diagnostic area under the receiver-operating characteristic curve of 0.84 (Fig. 3d).

Discussion

The early detection of cancer from liquid biopsy-based diagnostics holds great potential; however, it is currently limited as a result of a low signal-to-noise ratio. ctDNA-based diagnostics have benefited from consideration of information beyond the primary sequence, particularly methylation data. Although there are reported associations between RNA epitranscriptomic modifications and cancer, their challenging quantification has proved a barrier to incorporation into multi-omic liquid biopsies. We present here a Nanopore-based method that, when coupled to AI algorithms, can accurately profile the RNA modification profile of small RNA biomarkers from blood, and furthermore demonstrate the diagnostic utility of a differentially Nm methylated rRF in lung cancer liquid biopsy samples.

The majority of currently developed cancer screening liquid biopsies are focused on the detection of nucleic acid biomarkers; however, the greatest source of cell-free circulating nucleic acids is apoptotic haematopoietic cells[46]. This creates a vast background signal of both cfDNA and cfRNA in comparison to the relatively rare nucleic acid biomarkers that have been released from the tumor. This is a particular challenge in the early stages of disease, when the tumor is small and its rate of shedding relatively minute[47]. This has led to the search for additional features that may help to differentiate a true cancer-derived biomarker from the background signal. Unique cancer mutations present in ctDNA can be exploited; however, these may fall under the limit of detection, be the result of sequencing errors[48], or even false positive signals derived from clonal haematopoiesis[49]. Searching deeper, epigenetic changes associated with cancer, including differential CpG methylation patterns[9,10], or altered fragmentation profiles resulting from changes in chromatin structure[11], have fruitfully improved the performance of liquid biopsy diagnostics. The study of equivalent epitranscriptomic modifications and their biological function is less mature, although several have been noted in association with cancer. Unfortunately, current tools for their quantification, including immunoprecipitation and mass spectrometry-based methods, are cumbersome and pose a barrier to incorporation into clinical liquid biopsy diagnostics.

Third-generation sequencing technologies, including ONT and Pacific Biosciences, are unique in their ability to simultaneously analyse both sequence and chemical modifications within a nucleic acid. We are already seeing the application of this technology to the dual characterisation of both ctDNA sequence and methylation for liquid biopsies[50]. ONT technology has also shown promise in the analysis of cfRNA liquid biopsy samples[5]; however, current implementations have only sequenced cDNA, which

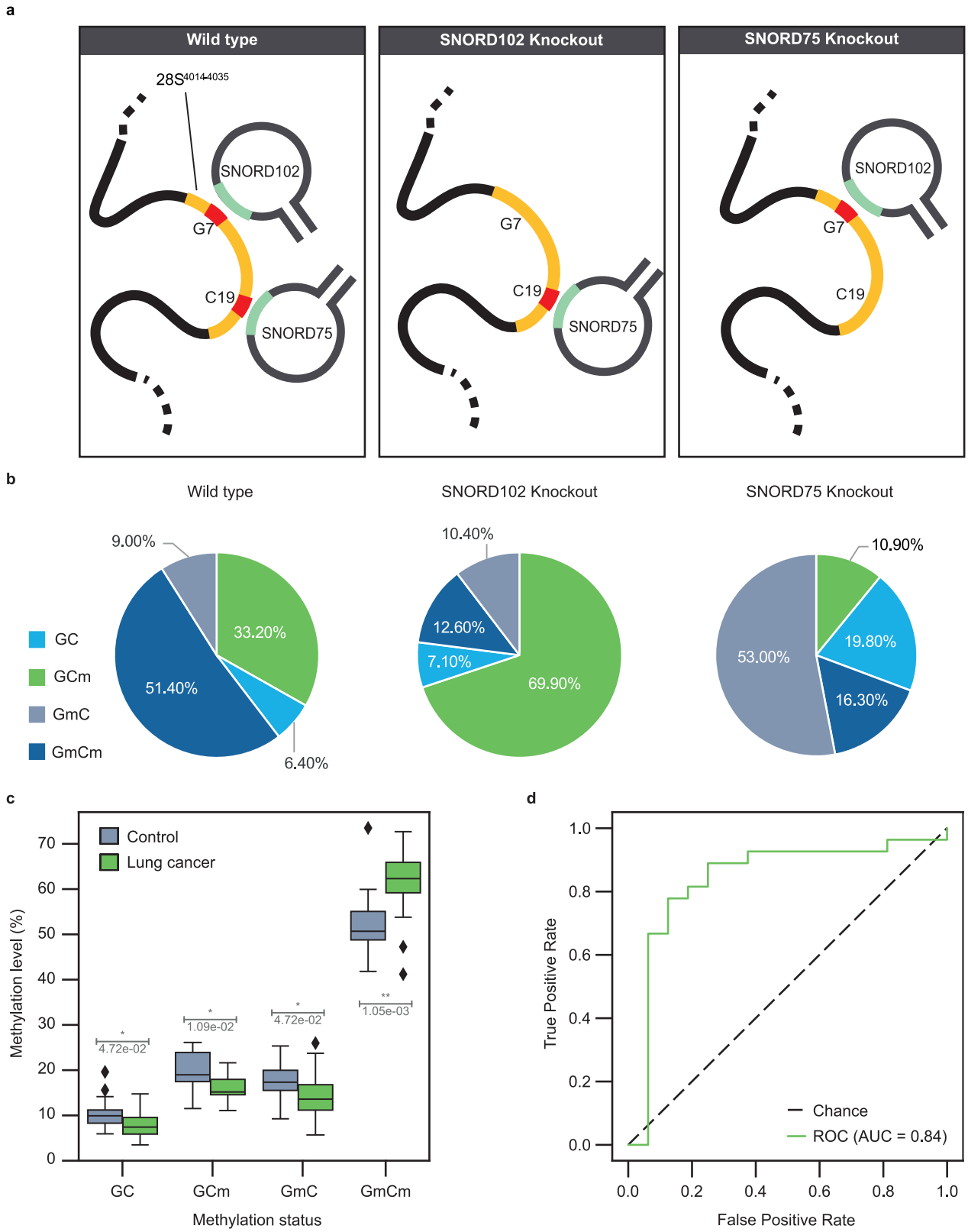

Fig. 3 | 28S$^{4014–4035}$ methylation dynamics in a cellular model and liquid biopsy samples from lung cancer patients. a Schematic representation of the templated deposition of Nm modifications on the rRNA directed by SNORDs. The 28S$^{4014–4035}$ fragment is marked in orange, and SNORD-directed Nm in red. **b** The proportion 28S$^{4014–4035}$ in each of its four possible Nm methylation states was measured from conditioned growth media of wild-type, SNORD102, or SNORD75 knockout HCC-827 cells, respectively. **c** Box and whisker plot quantifying the proportion of each 28S$^{4014–4035}$ methylation state from cfRNA samples collected from control ($n = 16$) and lung cancer patients ($n = 27$). Significance testing was performed using a two-tailed t-test. The t-statistics are as follows: GC: −2.33, GmC: −2.38, GCm: −3.23, GmCm: 4.14. p values corrected for multiple testing with the Holm method are displayed in the figure. For each methylation status and group, we plot a box using the first and third quartiles and the median line. Outliers are determined by 1.5× the IQR, and the whiskers show the remaining domain of values. **d** Receiver-operating characteristic curve using the proportion of GmCm 28S$^{4014–4035}$ as a biomarker for lung cancer.

Table 1 | Clinical characteristics of the patient cohort

Patient characteristic	Controls (n = 16)	Cancers (n = 27)
Sex, *n* (%)		
Female	8 (50.0)	5 (18.5)
Male	8 (50.0)	22 (81.5)
Age, years		
Mean ± SD	62.4 ± 8.2	68.0 ± 7.2
Median (range)	63 (50–79)	69 (53–79)
Smoking status, *n* (%)		
Current	12 (75.0)	21 (77.8)
Former	4 (25.0)	6 (22.2)
Pack-years, mean ± SD	39.2 ± 11.0	54.0 ± 15.5
Histologic subtype, *n* (%)		
NSCLC adenocarcinoma		5 (18.5)
NSCLC squamous		14 (51.9)
NSCLC other		3 (11.1)
SCLC		5 (18.5)
Pathologic stage, *n* (%)		
Ia		1 (3.8)
Ib		1 (3.8)
IIb		1 (3.8)
IIIa		2 (7.7)
IIIb		4 (15.4)
IIIc		2 (7.7)
IVa		6 (23.1)
IVb		8 (30.8)
No information		2 (7.7)

therefore preclude the investigation of the impact of potential RNA modifications that are lost in the process of reverse transcription during library preparation. A growing body of research is now harnessing the data measured during DRS to capture RNA sequence and epitranscriptomic modification profile at the single-molecule resolution[25–29]. The method we have developed here exploits this by using DRS to sequence a cancer biomarker in liquid biopsy samples and enables the streamlined incorporation of RNA modification information to diagnostic tests.

The biological role for the cancer-associated differential methylation of $28S^{4014–4035}$ is at present unclear. It has been reported that Nm modifications may have a role in regulating the translational efficiency of the ribosome[51]. Furthermore, the SNORD-directed methyltransferase Fibrillarin has been shown to be upregulated in the background of inactivated p53, leading to global alterations in rRNA Nm profiles and increased translational efficiency of specific transcripts, including that of the oncogene cMyc[52]. This raises the question as to whether the increased proportion of $28S^{4014–4035}$ displaying a GmCm pattern in the plasma of lung cancer patients directly reflects an oncogenic process or is just a bystander. Further understanding of the role of RNA modifications in the ribosome and its fragmentation into rRFs may lead to the identification of more novel cancer biomarkers.

Limitations of this work include the use of RNA002 DRS chemistry, which has since been replaced by RNA004 and whose use would necessitate the retraining of the AI classification algorithms presented here. Furthermore, the low efficiency of targeted RNA capture in the library preparation from biological samples currently limits the method to highly expressed small RNA species such as $28S^{4014–4035}$. This could theoretically be overcome by enrichment techniques such as ultracentrifugation for exosomal RNAs or immunoprecipitation of protein-bound RNAs, for example, miRNAs contained within Ago2. A further limitation is the focus on the thorough characterisation of a single small RNA cancer biomarker with known Nm methylation sites. We envision that this method will be readily applicable to other small RNAs with known epitranscriptomic modifications through adaptation of 5′ and 3′ adapter design, and AI algorithm retraining. Most exciting is the prospect that advances in RNA modification-aware base-calling algorithms may enable the unbiased analysis of RNA modification across the small RNA transcriptome.

Together, we have developed a method for the targeted capture of small RNA biomarkers into a DRS-compatible sequencing library, trained AI algorithms to accurately profile epitranscriptomic modifications at the single-molecule level, and discovered differential methylation of a lung cancer biomarker in liquid biopsy samples. We hope that this tool will enable the greater exploration of epitranscriptomic information that may be integrated alongside other omics methods for improved liquid biopsy diagnostic performance.

Data availability

The raw Nanopore fast5 files for all training datasets have been deposited online in the ENA (PRJEB86051). The source data for Figs. 2e and 3b–d can be found in the Supplementary Data.

Code availability

The code required to train, test, and infer barcode and methylation status from Nanopore data is available on GitHub (https://github.com/gitHBDX/Nano-smallRNAseq).

Received: 22 August 2024; Accepted: 31 October 2025;
Published online: 16 December 2025

References

1. Bettegowda, C. et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6**, 224ra24 (2014).
2. Phallen, J. et al. Direct detection of early-stage cancers using circulating tumor DNA. *Sci. Transl. Med*. **9**, eaan2415 (2017).
3. Fish, L. et al. Cancer cells exploit an orphan RNA to drive metastatic progression. *Nat. Med.* **24**, 1743–1751 (2018).
4. Larson, M. H. et al. A comprehensive characterization of the cell-free transcriptome reveals tissue- and subtype-specific biomarkers for cancer detection. *Nat. Commun.* **12**, 2357 (2021).
5. Reggiardo, R. E. et al. Profiling of repetitive RNA sequences in the blood plasma of patients with cancer. *Nat. Biomed. Eng.* **7**, 1627–1635 (2023).
6. Dnistrian, A. M., Schwartz, M. K., Greenberg, E. J., Smith, C. A. & Schwartz, D. C. Evaluation of CA M26, CA M29, CA 15-3 and CEA as circulating tumor markers in breast cancer patients. *Tumor Biol.* **12**, 82–90 (1991).
7. Crawford, L. V., Pim, D. C. & Bulbrook, R. D. Detection of antibodies against the cellular protein p53 in sera from patients with breast cancer. *Int. J. Cancer* **30**, 403–408 (1982).
8. Ben-Mahrez, K. et al. Circulating antibodies against c-MYC oncogene product in sera of colorectal cancer patients. *Int. J. Cancer* **46**, 35–38 (1990).
9. Shen, S. Y. et al. Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature* **563**, 579–583 (2018).
10. Liu, M. C. et al. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann. Oncol.* **31**, 745–759 (2020).
11. Cristiano, S. et al. Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* **570**, 385–389 (2019).
12. Zhou, Z. et al. Fragmentation landscape of cell-free DNA revealed by deconvolutional analysis of end motifs. *Proc. Natl. Acad. Sci. USA* **120**, e2220982120 (2023).
13. Konno, M. et al. Distinct methylation levels of mature microRNAs in gastrointestinal cancers. *Nat Commun* **10**, 3888 (2019).
14. Seok, H. et al. Position-specific oxidation of miR-1 encodes cardiac hypertrophy. *Nature* **584**, 279–285 (2020).

15. Liang, H. et al. 3′-Terminal 2′-O-methylation of lung cancer miR-21-5p enhances its stability and association with Argonaute 2. *Nucleic Acids Res.* **48**, gkaa504 (2020).
16. Wang, H., Song, X., Song, C., Wang, X. & Cao, H. m6A-seq analysis of microRNAs reveals that the N6-methyladenosine modification of miR-21–5p affects its target expression. *Arch. Biochem. Biophys.* **711**, 109023 (2021).
17. Nachtergaele, S. & He, C. Chemical Modifications in the Life of an mRNA Transcript. *Annu. Rev. Genet.* **52**, 1–24 (2018).
18. Jain, M., Abu-Shumays, R., Olsen, H. E. & Akeson, M. Advances in nanopore direct RNA sequencing. *Nat. Methods* **19**, 1160–1164 (2022).
19. Workman, R. E. et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* **16**, 1297–1305 (2019).
20. Diensthuber, G. et al. Enhanced detection of RNA modifications and read mapping with high-accuracy nanopore RNA basecalling models. *Genome Res.* **34**, 1865–1877 (2024).
21. Leger, A. et al. RNA modifications detection by comparative Nanopore direct RNA sequencing. *Nat. Commun.* **12**, 7198 (2021).
22. Pratanwanich, P. N. et al. Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. *Nat. Biotechnol.* **39**, 1394–1402 (2021).
23. Abebe, J. S. et al. DRUMMER—rapid detection of RNA modifications through comparative nanopore sequencing. *Bioinformatics* **38**, btac274 (2022).
24. Hassan, D., Ariyur, A., Daulatabad, S. V., Mir, Q. & Janga, S. C. Nm-Nano: a machine learning framework for transcriptome-wide single-molecule mapping of 2´-O-methylation (Nm) sites in nanopore direct RNA sequencing datasets. *RNA Biol.* **21**, 560–574 (2024).
25. Hendra, C. et al. Detection of m6A from direct RNA sequencing using a multiple instance learning framework. *Nat. Methods* **19**, 1590–1598 (2022).
26. Cruciani, S. et al. De novo basecalling of RNA modifications at single molecule and nucleotide resolution. *Genome Biol.* **26**, 38 (2025).
27. Mateos, P. A. et al. Prediction of m6A and m5C at single-molecule resolution reveals a transcriptome-wide co-occurrence of RNA modifications. *Nat. Commun.* **15**, 3899 (2024).
28. Wu, Y. et al. Transfer learning enables identification of multiple types of RNA modifications using nanopore direct RNA sequencing. *Nat. Commun.* **15**, 4049 (2024).
29. Chan, A., Vries, I. S. N., Scheitl, C. P. M., Höbartner, C. & Dieterich, C. Detecting m6A at single-molecular resolution via direct RNA sequencing and realistic training data. *Nat. Commun.* **15**, 3323 (2024).
30. Krogh, N. et al. Profiling of 2′-O-Me in human rRNA reveals a subset of fractionally modified positions and provides evidence for ribosome heterogeneity. *Nucleic Acids Res.* **44**, 7884–7895 (2016).
31. Sharma, S., Marchand, V., Motorin, Y. & Lafontaine, D. L. J. Identification of sites of 2′-O-methylation vulnerability in human ribosomal RNAs by systematic mapping. *Sci. Rep.* **7**, 11490 (2017).
32. Taoka, M. et al. Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res.* **46**, gky811 (2018).
33. Guan, L. & Grigoriev, A. Computational meta-analysis of ribosomal RNA fragments: potential targets and interaction mechanisms. *Nucleic Acids Res.* **49**, 4085–4103 (2021).
34. Sikosek, T. et al. Early detection of lung cancer using small RNAs. *J. Thorac. Oncol*. https://doi.org/10.1016/j.jtho.2023.07.005 (2023).
35. Krause, M. Flongle DirectRNA Library preparation v1. https://doi.org/10.17504/protocols.io.bcwcixaw (2020).
36. Lucas, M. C. et al. Quantitative analysis of tRNA abundance and modifications by nanopore RNA sequencing. *Nat. Biotechnol*. 1–15. https://doi.org/10.1038/s41587-023-01743-6 (2023).
37. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at *arXiv* https://doi.org/10.48550/arxiv.1303.3997 (2013).
38. Truong, C., Oudre, L. & Vayatis, N. Selective review of offline change point detection methods. *Signal Process.* **167**, 107299 (2020).
39. Bao, Y. et al. SquiggleNet: real-time, direct classification of nanopore signals. *Genome Biol.* **22**, 298 (2021).
40. Vaclova, T. et al. Clinical impact of subclonal EGFR T790M mutations in advanced-stage EGFR-mutant non-small-cell lung cancers. *Nat. Commun.* **12**, 1780 (2021).
41. Garalde, D. R. et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* **15**, 201–206 (2018).
42. Saville, L. et al. NERD-seq: a novel approach of Nanopore direct RNA sequencing that expands representation of non-coding RNAs. *Genome Biol.* **25**, 233 (2024).
43. Thomas, N. K. et al. Direct Nanopore sequencing of individual full length tRNA strands. *ACS Nano* **15**, 16642–16653 (2021).
44. Honda, S. & Kirino, Y. Dumbbell-PCR: a method to quantify specific small RNA variants with a single nucleotide resolution at terminal sequences. *Nucleic Acids Res.* **43**, e77–e77 (2015).
45. Smith, M. A. et al. Molecular barcoding of native RNAs using nanopore sequencing and deep learning. *Genome Res.* **30**, 1345–1353 (2020).
46. Lui, Y. Y. et al. Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. *Clin. Chem.* **48**, 421–427 (2002).
47. Avanzini, S. et al. A mathematical model of ctDNA shedding predicts tumor detection size. *Sci. Adv.* **6**, eabc4308 (2020).
48. Ma, X. et al. Analysis of error profiles in deep next-generation sequencing data. *Genome Biol.* **20**, 50 (2019).
49. Chan, H. T., Chin, Y. M., Nakamura, Y. & Low, S.-K. Clonal hematopoiesis in liquid biopsy: from biological noise to valuable clinical implications. *Cancers* **12**, 2277 (2020).
50. Lau, B. T. et al. Single-molecule methylation profiles of cell-free DNA in cancer with nanopore sequencing. *Genome Med.* **15**, 33 (2023).
51. Erales, J. et al. Evidence for rRNA 2′-O-methylation plasticity: Control of intrinsic translational capabilities of human ribosomes. *Proc. Natl. Acad. Sci. USA* **114**, 12934–12939 (2017).
52. Marcel, V. et al. p53 acts as a safeguard of translational control by regulating fibrillarin and rRNA methylation in cancer. *Cancer Cell* **24**, 318–330 (2013).

Acknowledgements

The authors would like to thank the clinical team at FnSP, particularly Jana Maličká, as well as the Hummingbird Team, including Patrick Neubert, Jurgita Juozaityte, Elena Gleim, Elena Neumüller, Kathrin Waurich, Lotta Simon, and, not least, Jochen Kohlhaas. The authors thank Diana Ordonez from EMBL for assistance with the cell sorting, and Vladimir Benes and Laura Villacorta from EMBL for discussions and advice regarding Nanopore sequencing. Finally, the authors would like to thank all study participants and their families. Hummingbird Diagnostics GmbH funded this study.

Author contributions

M.S.D., M.F., and T.R. designed and planned the study. T.Š., F.H., C.R., R.W., K.T., and M.U. organised clinical enrollment and sample, and data collection. M.S.D., A.D.M., E.M., J.C., J.S., and C.B.S. processed biological material and performed experiments. M.S.D., M.F., M.K., T.S., B.R.S., R.H., and T.R. analysed and interpreted the data. M.S.D., M.F., B.R.S., and T.R. wrote the manuscript. All authors approved the final version.

Competing interests

M.S.D., M.F., M.K., A.D.M., E.M., J.C., J.S., C.B.S., F.H., C.R., R.W., K.T., T.S., B.R.S., R.H., and T.R. are employees of Hummingbird Diagnostics and hold company stock options. M.S.D., M.F., M.K., B.R.S., R.H., and T.R. are inventors of patent applications related to Nanopore RNA modification profiling-based diagnostics submitted by Hummingbird Diagnostics. The remaining authors declare no competing interests.

Consent for publication

All patients and authors consented to the publication of the results.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43856-025-01235-5.

Correspondence and requests for materials should be addressed to Timothy Rajakumar.

Peer review information *Communications Medicine* thanks Xiang Yu, Quentin Gouil, Atocha Romero, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025