



Identification of nucleoside monophosphates and their epigenetic modifications using an engineered nanopore

Yuqin Wang^{1,2,3}, Shanyu Zhang^{1,2,3}, Wendong Jia^{1,2,3}, Pingping Fan^{1,2}, Liying Wang^{1,2}, Xinyue Li^{1,2}, Jialu Chen^{1,2}, Zhenyuan Cao^{1,2}, Xiaoyu Du^{1,2}, Yao Liu^{1,2}, Kefan Wang^{1,2}, Chengzhen Hu^{1,2}, Jinyue Zhang^{1,2}, Jun Hu¹, Panke Zhang¹, Hong-Yuan Chen¹ and Shuo Huang^{1,2} ✉

RNA modifications play critical roles in the regulation of various biological processes and are associated with many human diseases. Direct identification of RNA modifications by sequencing remains challenging, however. Nanopore sequencing is promising, but the current strategy is complicated by sequence decoding. Sequential nanopore identification of enzymatically cleaved nucleoside monophosphates may simultaneously provide accurate sequence and modification information. Here we show a phenylboronic acid-modified hetero-octameric *Mycobacterium smegmatis* porin A nanopore, with which direct distinguishing between monophosphates of canonical nucleosides, 5-methylcytidine, *N*⁶-methyladenosine, *N*⁷-methylguanosine, *N*¹-methyladenosine, inosine, pseudouridine and dihydrouridine was achieved. A custom machine learning algorithm, which reports an accuracy of 0.996, was also applied to the quantitative analysis of modifications in microRNA and natural transfer RNA. It is generally suitable for sensing of a variety of other nucleoside or nucleotide derivatives and may bring new insights to epigenetic RNA sequencing.

Many RNA modifications are enzymatically driven chemical modifications to either the ribose or the nucleobase of nucleotides. Approximately 170 types of RNA modifications are known¹ and are essential for various biological processes such as genetic recoding², pre-messenger RNA (mRNA) splicing³, mRNA exporting⁴, RNA folding⁵ and chromatin state regulation⁶. Accumulating evidence indicates that a large number of RNA modifications are associated with cancers⁷, neurological disorders⁸ and other human diseases⁹, and may thus be treated as either diagnostic markers or therapeutic targets. Recent reports also indicate that RNA modifications are associated with the yield of grains¹⁰. However, there is an unmet but urgent need to map diverse RNA modifications accurately, and this is complicated by the similarity in their chemical structures¹¹.

Analysis of RNA modifications can be performed by thin layer chromatography¹², high performance liquid chromatography coupled with UV spectrophotometry¹³ or high performance liquid chromatography coupled to mass spectrometry¹⁴. However, they all fail to provide any sequence information. Methods based on next-generation sequencing allow for mapping of transcriptome-wide RNA modifications¹⁵, but they rely on either specific antibodies¹⁶ or chemical treatments of RNA¹⁷. These methods are typically tailored to only one specific modification, and thus only a limited type of modifications can be detected by sequencing. These include pseudouridine (ψ)¹⁸, *N*⁶-methyladenosine (m⁶A)¹⁹, 5-methylcytidine (m⁵C)²⁰, *N*¹-methyladenosine (m¹A)²¹, *N*⁷-methylguanosine (m⁷G)²², 5-hydroxymethylcytosine²³, *N*^{6,2'}-*O*-dimethyladenosine¹⁶, *N*⁴-acetylcytidine²⁴ and A-to-I editing²⁵. Third-generation sequencing techniques, including methods developed by Pacific Biosciences or Oxford Nanopore Technologies, may

overcome these shortcomings²⁶. In Pacific Biosciences sequencing, RNA modifications are identified by the observation of time variation between base incorporations²⁷. On the other hand, nanopore sequencing provided by Oxford Nanopore Technologies reports RNA modifications by identifying variations in the ionic current²⁸ or the event dwell time²⁹. However, the nanopore strand sequencing strategy³⁰ still suffers from a low spatial resolution, which is even worse when the modified nucleotides are close neighbours³¹.

Sequencing RNA in an *exo*-sequencing manner, is a different strategy with which exonuclease-decomposed nucleotides can be sequentially read by a nanopore. However, this requires the existence of a high resolution nanopore that can unambiguously recognize all nucleotides and their major modifications. A cyclodextrin embedded α -haemolysin (α -HL)^{32,33} was previously reported to perform this task, but the results fail to show true discrimination between cytidine diphosphate and uridine diphosphate. Identification of RNA modifications was also not demonstrated³². This low resolution should result from the cylindrical lumen geometry of α -HL³⁴. Instead, *Mycobacterium smegmatis* porin A (MspA)³⁵, which is a conically shaped pore widely applied in nanopore sequencing³⁶, single molecule chemistry³⁷ and structure profiling of biomacromolecules³⁸, is more advantageous. Phenylboronic acid (PBA) is known to form covalent bonds reversibly with 1,2 or 1,3-diols³⁹. Previously, the introduction of PBA to the nanopore lumen was successfully applied to the detection of various *cis*-diol-containing analytes such as saccharides⁴⁰, epinephrine and Remdesivir⁴¹. However, a hetero-octameric MspA nanopore containing a single PBA adaptor has not been reported previously and nanopore identification of a large variety of epigenetically modified nucleoside monophosphates (NMPs) has also never been reported.

¹State Key Laboratory of Analytical Chemistry for Life Sciences, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing, China.

²Chemistry and Biomedicine Innovation Center (ChemBIC), Nanjing University, Nanjing, China. ³These authors contributed equally: Yuqin Wang, Shanyu Zhang, Wendong Jia. ✉e-mail: shuo.huang@nju.edu.cn

NMP identification using a PBA-modified MspA

To build a hetero-octameric MspA, two different genes coding for N90C MspA-H6 and M2 MspA-D16H6, respectively, (Supplementary Table 1) were custom-synthesized. Both genes were simultaneously inserted into a pETDuet-1 co-expression vector (Methods). Specifically, the N90C MspA-H6 codes for an MspA monomer, at the pore constriction in which a sole cysteine is placed, whereas the M2 MspA-D16H6 codes for the monomer that does not contain any cysteine. Hetero-octameric MspAs composed of different fractions of both gene expression products were generated by prokaryotic co-expression (Supplementary Fig. 1) and were characterized by gel electrophoresis (Supplementary Figs. 2 and 3). The hetero-octameric MspA consisting of one unit of N90C MspA-H6 and seven units of M2 MspA-D16H6 is the only desired MspA assembly and is referred to as $(N90C)_1(M2)_7$ (Fig. 1a). $(N90C)_1(M2)_7$ was separated from other MspA hetero-octamers by high resolution gel electrophoresis followed with gel extraction (Methods, Supplementary Figs. 2 and 3). Subsequently, 3-(maleimide) phenylboronic acid (MPBA) was allowed to react with the sole cysteine of $(N90C)_1(M2)_7$ (Fig. 1b). A real-time observation of this reaction at the level of a single molecule was carried out by single channel recording in a 1.5 M KCl, 10 mM MOPS, pH 7.0 buffer (Fig. 1c and Methods). With a single $(N90C)_1(M2)_7$ inserted in the membrane and a continually applied +200 mV bias, the open pore current of $(N90C)_1(M2)_7$ (I_o) measures ~620 pA. Upward noises, which result from the cysteine residue at the pore constriction as previously reported⁴², were also observed. With the addition of MPBA to *cis* at a final concentration of 1 mM, a single current drop measuring ~100 pA was immediately observed. The previously observed upward noises also disappeared simultaneously, suggesting that the cysteine residue has been occupied and the PBA modification to the pore constriction was successful. For simplicity, this PBA-modified MspA is referred to as MspA-PBA. Under the same conditions, the open pore current of MspA-PBA (I_p) measures ~520 pA (Fig. 1c). MspA-PBA can also be prepared in ensemble by mixing $(N90C)_1(M2)_7$ with MPBA (Methods). If not otherwise stated, all subsequent measurements were carried out using ensemble-prepared MspA-PBA (Supplementary Fig. 4). Statistical results of the open pore current of $(N90C)_1(M2)_7$ and MspA-PBA are measured at 623 ± 13 (mean \pm full width half maximum (FWHM)) pA and 510 ± 14 (mean \pm FWHM) pA (Supplementary Fig. 4), consistent with the results previously measured (Fig. 1c). *I*-*V* curves of $(N90C)_1(M2)_7$ and MspA-PBA acquired with varying concentrations of KCl (0.15–2 M KCl) are presented in Supplementary Fig. 5. According to the slope of the *I*-*V* curves, the conductance of MspA-PBA measured with a 1.5 M KCl buffer was derived to be ~2.91 nS.

NMPs consist of a ribose, a phosphate group and a nucleobase, serving as monomeric units of RNA. Due to the presence of a *cis*-diol in the ribose, NMPs possess an affinity to PBA⁴³ and may be directly detected by MspA-PBA. To test this, single channel recording was performed using MspA-PBA in a 1.5 M KCl buffer (1.5 M KCl, 10 mM MOPS, pH 7.0) (Methods). A transmembrane potential of +200 mV was continually applied. Four canonical NMPs, adenine mononucleotide (AMP), guanine mononucleotide (GMP), cytosine mononucleotide (CMP) and uracil mononucleotide (UMP) were tested as analytes (Fig. 1d). Successive resistive pulses caused by NMPs were immediately observed (Fig. 1d). However, no events were observed when M2 MspA was tested, confirming that the PBA located at the pore constriction is critical in the generation of NMP sensing events (Supplementary Fig. 6). With MspA-PBA, deoxyribonucleoside monophosphate (dNMP) fails to report any events (Supplementary Fig. 7). This is expected because dNMPs have no *cis*-diol structure which is necessary for sensing.

To describe NMP sensing events quantitatively, the event dwell time (t_{off}), the interevent interval (t_{on}), the percentage blockage

($\%I_b = (I_p - I_b)/I_p$) and the noise amplitude (SD) were derived as described in Supplementary Fig. 8. Generally, the histograms of t_{off} and t_{on} show an exponential distribution, and could be fitted to derive the mean time constants τ_{off} or τ_{on} , respectively. The histograms of $\%I_b$ and SD show a Gaussian distribution, which could be fitted to derive the mean percentage blockage $\%I_b$ and SD, respectively. During NMP sensing, by varying the NMP concentrations in *cis*, the reciprocal of dwell time ($1/\tau_{\text{off}}$) remains constant. The reciprocal of the interevent interval ($1/\tau_{\text{on}}$), however, linearly correlates with the NMP concentration in *cis* (Supplementary Tables 2–5 and Supplementary Figs. 9–12). The dependence of the applied voltage during NMP sensing was also investigated using AMP as a representative analyte. Generally, when the voltage is upregulated, the $1/\tau_{\text{off}}$ decreases and the $1/\tau_{\text{on}}$ increases (Supplementary Table 6 and Supplementary Fig. 13). This is expected because in a pH 7.0 buffer, the NMP is negatively charged and the electrophoretic force can strongly regulate the binding rate.

The conical lumen of MspA provides an excellent resolution with which to distinguish between analytes with minor structural differences³⁷. Bindings of different NMPs to MspA-PBA result in highly distinguishable event features (Fig. 1d). This difference is more amplified at a higher applied voltage (Supplementary Fig. 14) and all subsequent measurements were carried out at a voltage of +200 mV, if not otherwise stated. In this condition, events generated by different NMPs form highly distinguishable populations in the scatter plot of $\%I_b$ versus SD (Fig. 1e). The histograms of $\%I_b$ of different NMP events also show fully separated Gaussian distributions (Fig. 1e and Supplementary Fig. 15), in which CMP ($\%I_b = 7.1 \pm 0.2\%$, $N = 3$, N represents the number of independent measurements), UMP ($\%I_b = 8.64 \pm 0.09\%$, $N = 3$), AMP ($\%I_b = 10.89 \pm 0.14\%$, $N = 3$) and GMP ($\%I_b = 11.8 \pm 0.2\%$, $N = 3$) are fully resolved (Supplementary Table 7 and Supplementary Figs. 16 and 17). More details of NMP binding kinetics are also summarized in Supplementary Table 7 and Supplementary Fig. 17. Simultaneous sensing of CMP, UMP, AMP and GMP using MspA-PBA was also performed (Fig. 1f and Supplementary Fig. 18), from which different NMP identities can be directly called based on their distinct blockage characteristics. To the best of our knowledge, nanopore discrimination between canonical NMPs without any overlaps in the event distribution has never been previously reported.

Distinguishing of epigenetic NMPs

According to the literature, ~170 epigenetic NMPs have been previously discovered¹. These epigenetic NMPs have extremely minor structural differences and pose a great challenge for direct identification. This challenge may be solved by directly monitoring event features of nanopore readouts when epigenetic NMPs are bound to an MspA-PBA. To test this, the same measurements were carried out by taking monophosphates of m^5C , m^6A , m^7G , m^1A , inosine (I), ψ and dihydrouridine (D) as the analyte. Due to a lack of commercially available model compounds, ψ (Supplementary Fig. 19) and D (Supplementary Fig. 20) were custom-synthesized and characterized by WuXi AppTec. These epigenetic NMPs have covered the common types of modification occurring with canonical NMPs such as methylation, deamination, isomerization and reduction. Their nucleobase components are demonstrated in Fig. 2a. As shown in Fig. 2a and Supplementary Fig. 21, events of epigenetic NMPs have significantly different blockage amplitudes. To demonstrate a full comparison between all NMPs being tested to date, the $\%I_b$ distribution for each NMP is shown in a violin plot, demonstrating that almost all NMPs are already distinguishable solely by analysis of their $\%I_b$, though the event distributions of UMP and m^5C still have some overlaps (Fig. 2b). The big variations in ψ and m^7G result from the detection of non-specific events away from the main population of events. They may result from impurities introduced during synthesis of the compound. However, these

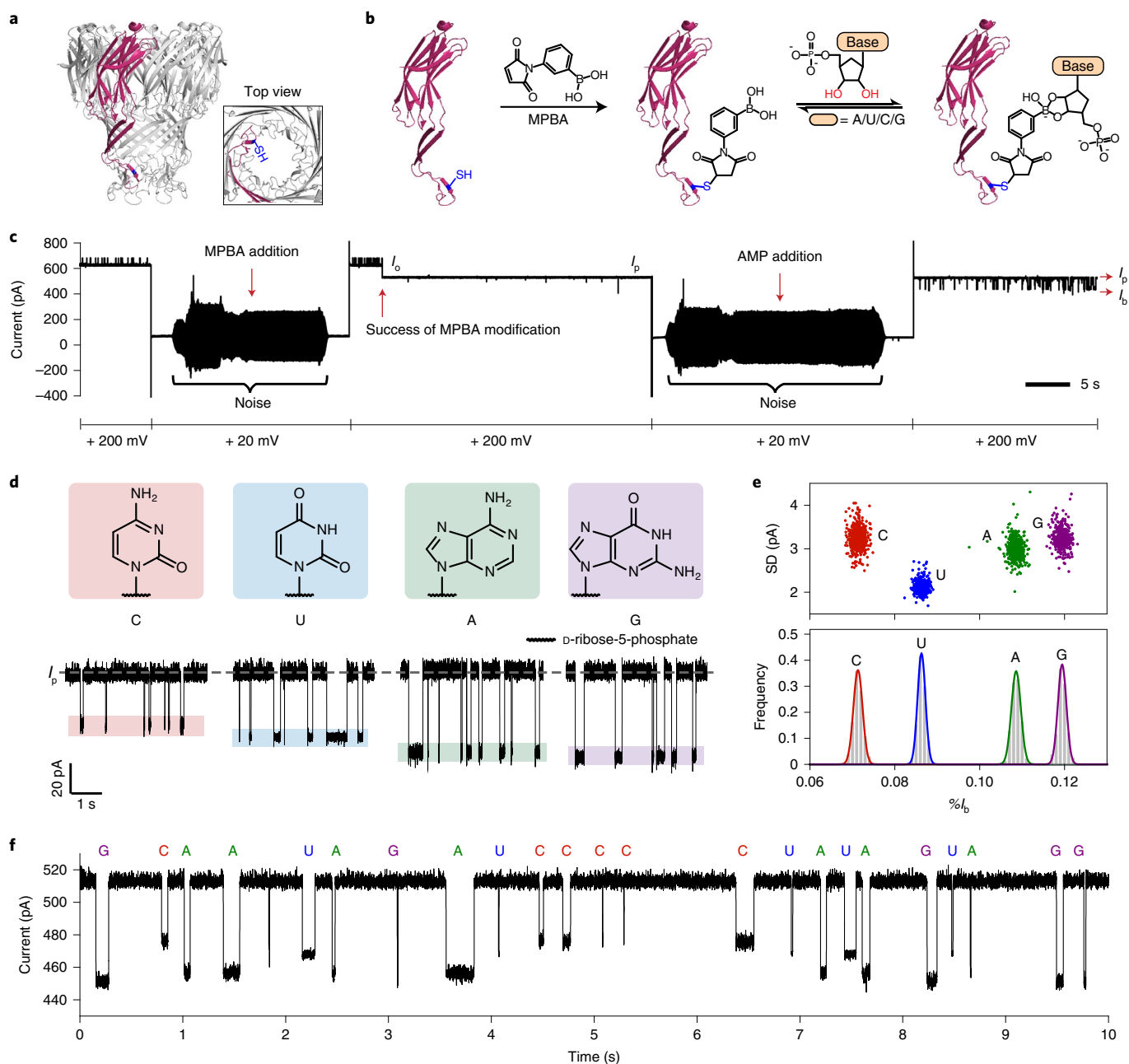


Fig. 1 | Discrimination of canonical NMPs using a PBA-modified MspA. **a**, The structure of $(N90C)_1(M2)_7$, $(N90C)_1(M2)_7$ is a hetero-octameric MspA composed of seven units of M2 MspA-D16H6 (grey) and one unit of N90C MspA-H6 (pink). $(N90C)_1(M2)_7$ contains a sole cysteine (blue). Square box: the top view of a $(N90C)_1(M2)_7$. **b**, The mechanism of NMP identification. A PBA was introduced to the pore constriction by modifying the sole cysteine thiol with a MPBA via Michael addition. NMPs can reversibly react with PBA, generating stochastic sensing events. **c**, Single channel observation of MPBA modification. After the addition of MPBA, a current drop of about 100 pA was observed, indicating the success of MPBA modification. With the subsequent addition of AMP, successive binding events immediately appear. To minimize bilayer rupture, the applied bias is switched to +20 mV whenever the Faraday cage is opened. The large noise is introduced during opening of the Faraday cage to perform MPBA or AMP addition. The open pore current (I_o) and the blockage level (I_b) are also marked. **d**, The NMPs and their corresponding events. Top: the chemical structures of CMP (C), UMP (U), AMP (A) and GMP (G), of which the nucleobases were clearly demonstrated. Bottom: representative sensing events. NMPs were added to *cis* with a final concentration of 300 μ M for each analyte. I_b is marked with a grey dashed line. The blockage levels are marked with colour bands. **e**, Top: a scatter plot of $\%I_b$ versus SD from results acquired with four types of NMPs. Bottom: corresponding event histogram of $\%I_b$. Events were acquired from four individual measurements, in which four types of NMP were separately added to *cis* with a final concentration of 300 μ M. Five-hundred successive events of each NMP were employed to generate the statistics. **f**, A representative trace when four types of NMPs were simultaneously sensed. NMPs were simultaneously added to *cis* with a final concentration of 300 μ M for each analyte. Events of different NMPs were identified according to the characteristic blockage depth and were marked, respectively, with C, U, A and G.

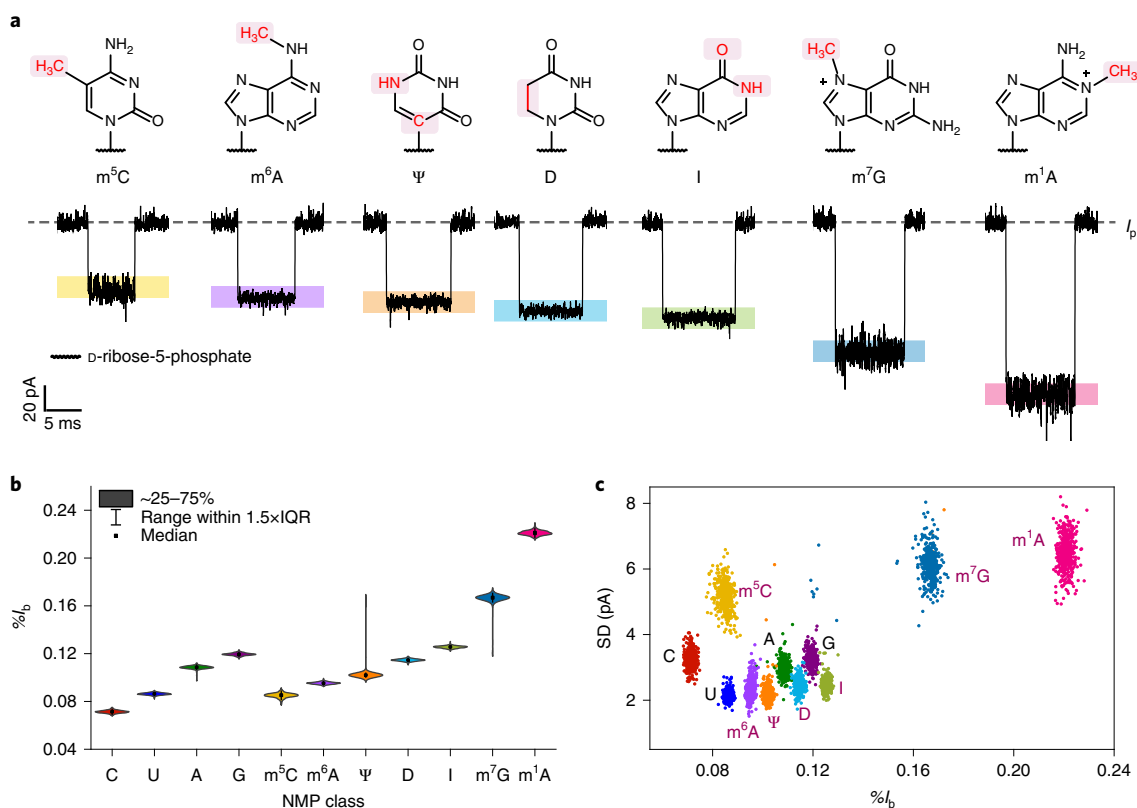


Fig. 2 | Epigenetic NMPs identified by MspA-PBA. **a**, Top: the epigenetic NMPs investigated in this paper. Seven types of epigenetic NMPs, including monophosphates of m^5C , m^6A , ψ , D, I, m^7G and m^1A were investigated. For ease of display, only the nucleobases are shown and all modifications are highlighted in red. Bottom: representative events of corresponding NMPs. From left to right, the representative events were, respectively, from m^5C , m^6A , ψ , D, I, m^7G and m^1A . The measurements were carried out as described in Methods in a 1.5 M KCl buffer (1.5 M KCl, 10 mM MOPS, pH 7.0). A transmembrane potential of +200 mV was continually applied. Epigenetic NMPs were added to *cis* with a final concentration of 300 μ M for each analyte. The open pore current (I_p) of MspA-PBA is marked with a dashed line. The blockage levels are marked with colour bands. Noticeable differences in $\%I_b$ and SD were observed. **b**, The violin plot of $\%I_b$ of different NMPs. Solely by analysing $\%I_b$, different NMPs are generally distinguishable except for U and m^5C . IQR, interquartile range. **c**, The scatter plot of $\%I_b$ versus SD for canonical and epigenetic NMPs. Events from 11 types of NMPs are clearly distinguishable when both $\%I_b$ and SD were considered. Events in **b** and **c** were acquired from 11 independent measurements, in which 11 types of NMP were separately added to *cis* as the sole analyte with a final concentration of 300 μ M. Five-hundred successive events of each NMP were employed to generate the statistics in **b** and **c**.

non-specific events only contribute to 0.9% and 1.7% of all events being detected, respectively (Supplementary Fig. 22). The noise characteristics of NMPs may also be included in event analysis to improve the discrimination performance (Supplementary Table 7 and Supplementary Figs. 23 and 24). By plotting a scatter plot containing $\%I_b$ versus SD of NMP sensing events acquired from 11 different analytes, 11 fully resolved event populations were generated, respectively corresponding to each NMP being sensed (Fig. 2c). This confirms that this sensing configuration is compatible with epigenetic NMPs and their events are fully distinguishable. Direct discrimination between these 11 types of NMPs using nanopores has never been reported before, however. The discrimination between epigenetic NMPs and their corresponding canonical counterparts is demonstrated in Supplementary Fig. 25.

NMP identification by machine learning

A machine learning algorithm was established to automatically identify NMPs. The overall training process includes dataset input, feature extraction and model building (Fig. 3a and Methods). All events in the dataset have known labels since they were acquired with a sole NMP with a known identity. The $\%I_b$ and SD of each event were automatically extracted using MATLAB to form a feature matrix. Main stream models were evaluated and they all demonstrated satisfactory validation accuracies, indicating that the input data are of a high

quality. Specifically, the Kernel Naïve Bayes model and linear Support Vector Machine (SVM) model reported the highest accuracy score of 0.996 (Supplementary Table 8). The Linear SVM model was selected based on its better performance with the testing set. The confusion matrix results based on model testing using the Linear SVM model are shown in Fig. 3b, in which most NMP sensing results report either 99% or 100% accuracy. In Fig. 3c, a decision boundary plot generated by the Linear SVM model is also demonstrated.

The previously trained Linear SVM model was employed to predict events with unknown identities. Modified NMPs were added to the *cis* side in the order of m^5C , m^6A , I, m^7G , m^1A , ψ and D with CMP, UMP, AMP and GMP already placed in *cis*. With the Linear SVM model, newly added NMPs can be accurately identified (Supplementary Figs. 26 and 27). To evaluate the training efficiency of the model, learning curves were generated respectively with training or validation data (Supplementary Fig. 28), from which it is conclusive that 176 events were required for the model to reach a 0.990 accuracy. To show event identification from a mixture, a representative trace containing events from 11 different NMPs is demonstrated in Fig. 3d and Supplementary Movie 1. Different NMP types can be recognized and the corresponding labels predicted by machine learning are marked above the trace. This efficiently assists automatic nanopore sensing of different NMPs in a real measurement scenario in which different NMPs exist as a mixture.

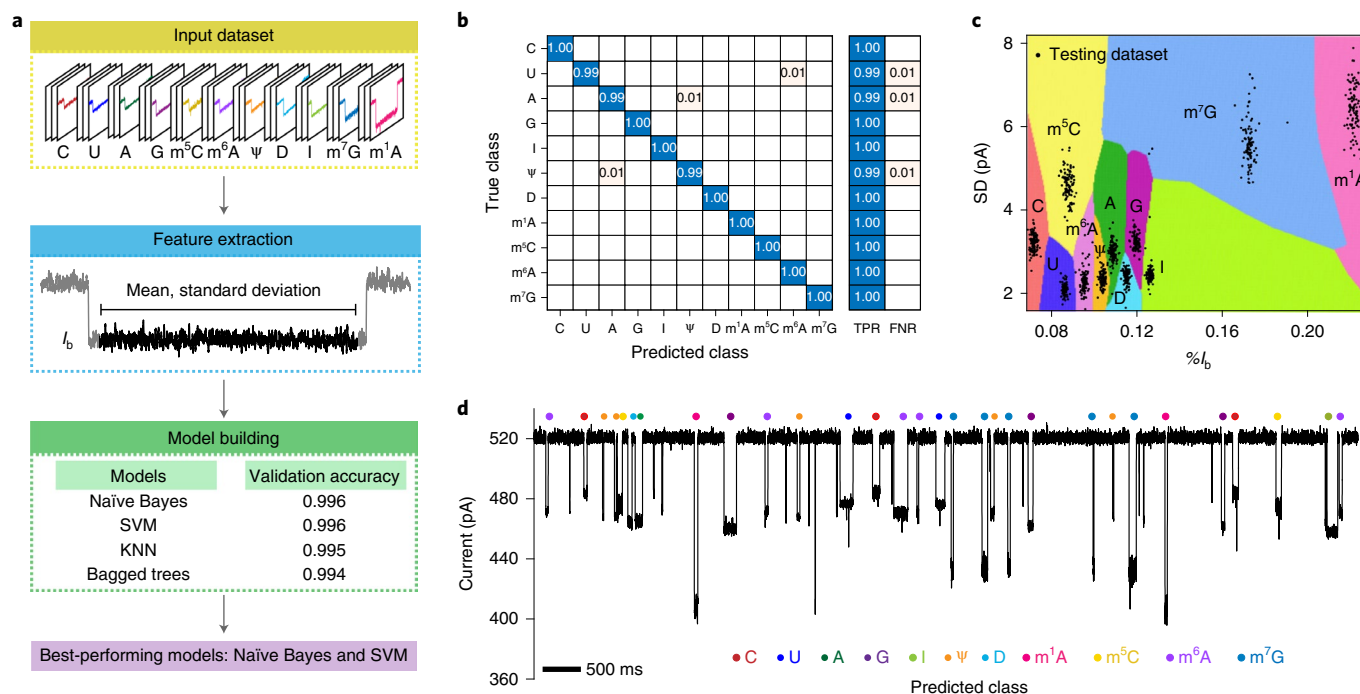


Fig. 3 | Machine learning assisted NMP identification. **a**, The flow diagram of the training process. Eleven classes of events, including C, U, A, G, m⁵C, m⁶A, ψ, D, I, m⁷G and m¹A were applied as the input dataset. Each class is composed of 400 events, randomly selected from a pool of events respectively acquired with each analyte type. The mean and the standard deviation of each event were extracted to form a feature matrix. Results in the matrix were further randomly split into a training subset for model training and a validation subset for model validation, with which a tenfold cross validation was performed. All classifiers in the Classification Learner toolbox of MATLAB, such as Naïve Bayes, SVM, K Nearest Neighbor (KNN) and Bagged trees, were evaluated to screen the best performing model. The SVM and the Naïve Bayes model have demonstrated the highest accuracy score of 0.996. The SVM was selected for all further investigations. **b**, The confusion matrix of NMP classification generated using the SVM model. One-hundred events from each NMP class were treated as the testing set. The true positive rate (TPR) and the false negative rate (FNR) were also demonstrated on the right. **c**, The decision boundary produced by the SVM model. Each coloured region represents the area in which a corresponding NMP event is to be predicted. **d**, A representative trace acquired by simultaneous sensing of 11 types of NMP. The measurements were carried out as described in Methods. The chambers were filled with a 1.5 M KCl buffer (1.5 M KCl, 10 mM MOPS, pH 7.0). A transmembrane potential of +200 mV was continually applied. NMPs were simultaneously added to *cis* with a final concentration of 100 μM for each analyte. Characteristic events from different NMPs were automatically predicted by the trained SVM models and labelled with different colour dots (CMP, red; UMP, blue; AMP, green; GMP, purple; m⁵C, yellow; m⁶A, orchid; ψ, orange; I, lime; D, cyan; m⁷G, teal; m¹A, pink).

Sensing of epigenetic NMPs from methylated microRNA

We further sought to demonstrate direct sensing of epigenetic NMPs in RNAs (Fig. 4a). By treatment with S1 nuclease, the RNA is first enzymatically decomposed into NMPs and then sensed by MspA-PBA. The observed nanopore events were identified by the previously trained machine learning model. Two microRNAs including hsa-miR-21 and hsa-miR-17 with known methylated sites⁴⁴ were applied. Specifically, the hsa-miR-21 contains a m⁵C at position 9 and the hsa-miR-17 contains a m⁶C at position 13 (Supplementary Table 9). Without any enzymatic treatment, hsa-miR-21 and hsa-miR-17 were sensed by MspA-PBA. However, only short-residing spiky events with undefined event amplitudes were observed (Supplementary Fig. 29), indicating that this sensing configuration is insensitive to the template RNAs itself. To minimize interferences of glycerol in the stock solution of S1 nuclease (Supplementary Fig. 30), the S1 nuclease was pretreated by ultrafiltration to remove glycerol (Methods and Supplementary Fig. 31). The pretreated S1 nuclease was then employed to digest the microRNAs at 23 °C for 4 h. From the gel electrophoresis results, both microRNAs were thoroughly decomposed (Methods and Supplementary Fig. 32). The enzymatic treatment product was then subjected to ultrafiltration to remove the S1 nuclease before nanopore measurements (Methods). During nanopore measurement, the hsa-miR-21 digestion product was added to *cis* with a final concentration

of 100 ng μl⁻¹. A representative trace is shown in Fig. 4b, in which many NMP binding events were observed, suggesting that the generated NMPs are detected well by MspA-PBA. The identities of NMPs were called by the algorithm, which are highly discriminable from the demonstrated NMP events (Supplementary Fig. 33).

According to the results acquired with hsa-miR-21, five types of NMPs were detected, including CMP, UMP, AMP, GMP and m⁵C (Fig. 4b,c), consistent with the hsa-miR-21 sequence composition (Supplementary Table 9). The abundance of each NMP type in hsa-miR-21 was also evaluated based on the rate of event appearance followed with a calibration (Method and Supplementary Table 10). The relative NMP composition in hsa-miR-21 was estimated to be 2.17 CMP, 6.81 UMP, 6.88 AMP, 4.92 GMP, 1.03 m⁵C, 0.06 I, 0.01 ψ and 0.10 D (Supplementary Fig. 34), generally consistent with the true values. The misjudgement of I, ψ and D result from the minor distribution overlap between AMP, ψ, GMP and I. However, the proportion of misjudgement is negligible. The feasibility of epigenetic NMP identification from miRNA is thus approved. To test its generality, hsa-miR-17, was also tested identically to that shown for hsa-miR-21. A representative trace containing nanopore sensing events of the digestion products of hsa-miR-17 is demonstrated in Fig. 4d. The scatter plot results demonstrate five dominant populations of NMP events, respectively corresponding to CMP, UMP, AMP, GMP and m⁶A (Fig. 4e), consistent with the sequence

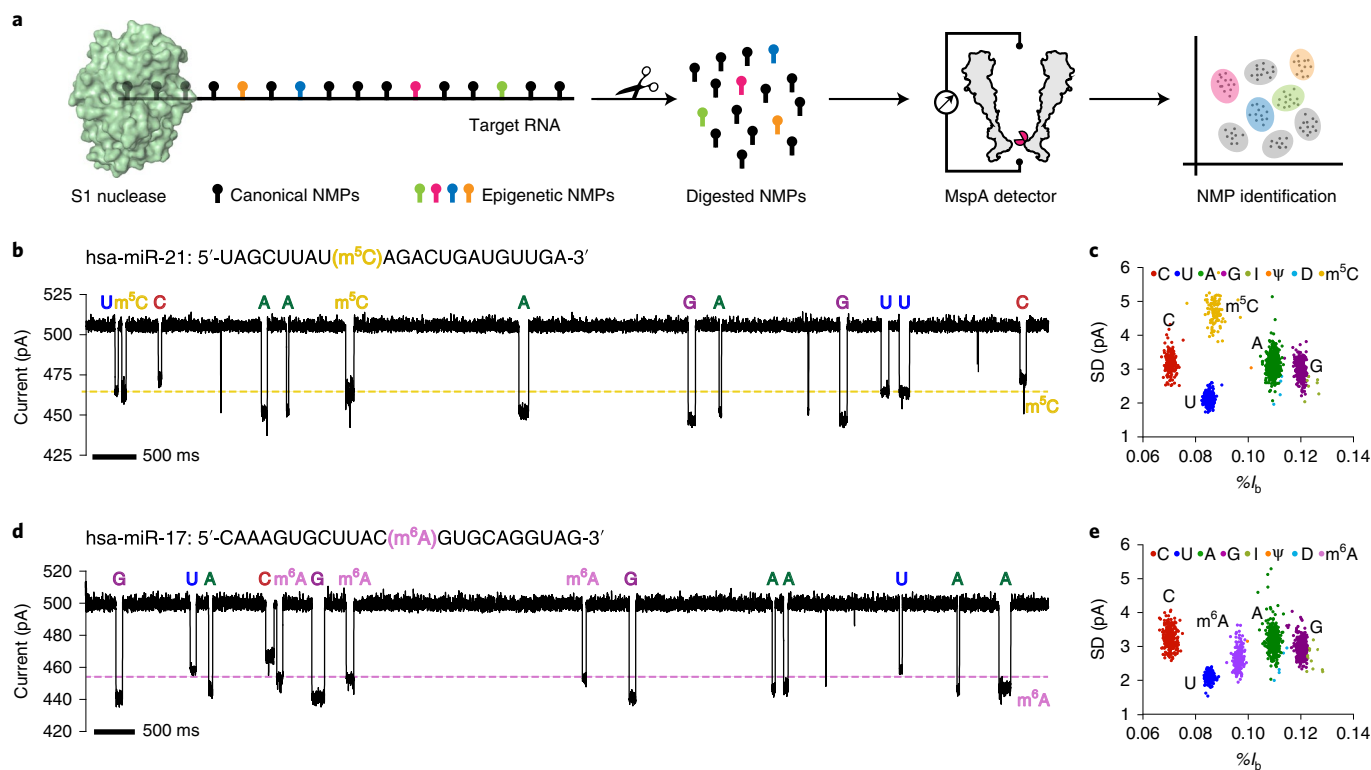


Fig. 4 | Detection of epigenetic modifications from RNA. **a**, The schematic diagram of NMP identification from RNA using MspA-PBA. S1 nuclease (green), an endonuclease insensitive to epigenetic modifications, was employed to decompose target RNAs into NMPs. The generated NMPs were then characterized using MspA-PBA, enabling profiling of RNA modifications in a quantitative manner. **b**, The sequence of hsa-miR-21 and the corresponding trace of nanopore sensing of the digested products. Hsa-miR-21 was reported to contain a m^5C at position 9. Characteristic events from C, U, A, G and m^5C were clearly detected in the trace, which are respectively marked. The blockage level of m^5C is marked with a yellow dashed line. Although the blockage level of m^5C is close to that of U, the noise of m^5C is significantly larger. **c**, The scatter plot of $\%I_b$ versus SD for hsa-miR-21 digestion products. Events demonstrated were acquired from a 60 min continuous recording. The NMP identity is predicted by the SVM model. Five populations, respectively, from events of C, U, A, G and m^5C were detected. **d**, The sequence of hsa-miR-17 and the corresponding trace of nanopore sensing of the digested products. Hsa-miR-17 was reported to contain a m^6A at position 13. Characteristic events of C, U, A, G and m^6A are clearly detected, and are marked with the corresponding labels. The blockage level of m^6A was marked with an orchid dashed line. **e**, The scatter plot of $\%I_b$ versus SD derived from sensing events of hsa-miR-17 digestion products. Events demonstrated were acquired from a 60 min continuous recording. The NMP identity was predicted by the SVM model. Five populations including C, U, A, G and m^6A were detected. All measurements were carried out as described in Methods in a 1.5 M KCl buffer (1.5 M KCl, 10 mM MOPS, pH 7.0). A transmembrane potential of +200 mV was continually applied. MicroRNA digestion product was added to *cis* with a final concentration of $100 \text{ ng } \mu\text{l}^{-1}$.

component of hsa-miR-17 (Supplementary Table 9). Quantitative analysis shows that the relative count of m^6A site is 1.08, indicating that only one m^6A site was present in the hsa-miR-17 (Supplementary Fig. 34), also consistent with expectations.

Detection of epigenetic NMPs from brewer's yeast tRNA^{Phe}

Transfer RNA (tRNA) is a type of low molecular weight RNA serving to link the mRNA sequence into the amino acid sequence of protein. Mature tRNAs also contain rich chemical modifications. As reported, more than 90 types of modifications have been discovered in tRNA⁴⁵. It is thus an ideal RNA to evaluate the performance of MspA-PBA in the identification of epigenetic modifications of natural samples. The brewer's yeast phenylalanine-specific tRNA (yeast tRNA^{Phe})³⁸ is applied as a model RNA to test its feasibility. As reported, a mature yeast tRNA^{Phe} contains 14 epigenetically modified sites originated from 11 types of modifications including N^2 -methylguanosine (m^2G), dihydrouridine (D), N^2,N^2 -dimethylguanosine (m^2_2G), 2'-*O*-methylcytidine (C_m), 2'-*O*-methylguanosine (G_m), wybutosine (Y), ψ , m^5C , m^7G , 5-methyluridine (T) and m^1A (Fig. 5a)⁴⁶. When the yeast tRNA^{Phe} is enzymatically decomposed into NMPs, monophosphates of D,

ψ , m^5C , m^7G , m^1A , m^2G , m^2_2G , T and Y are in principle detectable by MspA-PBA because their *cis-diol* structures remain unmodified. The event parameters of D, ψ , m^5C , m^7G and m^1A have been previously acquired and used for model training (Figs. 2a and Fig. 3) so that their events are identifiable by the machine learning algorithm. The monophosphates of m^2G , m^2_2G , T and Y are in principle detectable by MspA-PBA and new clusters of events are expected to be observed. However, due to a lack of corresponding pure compounds to produce events for training, the corresponding nanopore events are detectable but not identifiable. C_m and G_m , which lack a *cis-diol*, are in principle undetectable by MspA-PBA.

tRNA^{Phe} was first enzymatically treated with S1 nuclease at 23 °C for 15 h to produce NMPs (Methods). According to the gel electrophoresis result, it was confirmed that the tRNA^{Phe} had been thoroughly decomposed (Fig. 5b). The enzymatic treatment product was then ultra-filtrated to remove the S1 nuclease and used in subsequent nanopore measurements (Methods). Nanopore measurements were carried out with MspA-PBA (Methods). The yeast tRNA^{Phe} digestion product was added to *cis* with a final concentration of $100 \text{ ng } \mu\text{l}^{-1}$. The acquired raw events were shown in a scatter plot (Supplementary Fig. 35). Glycerol events, which were introduced by the stock solution

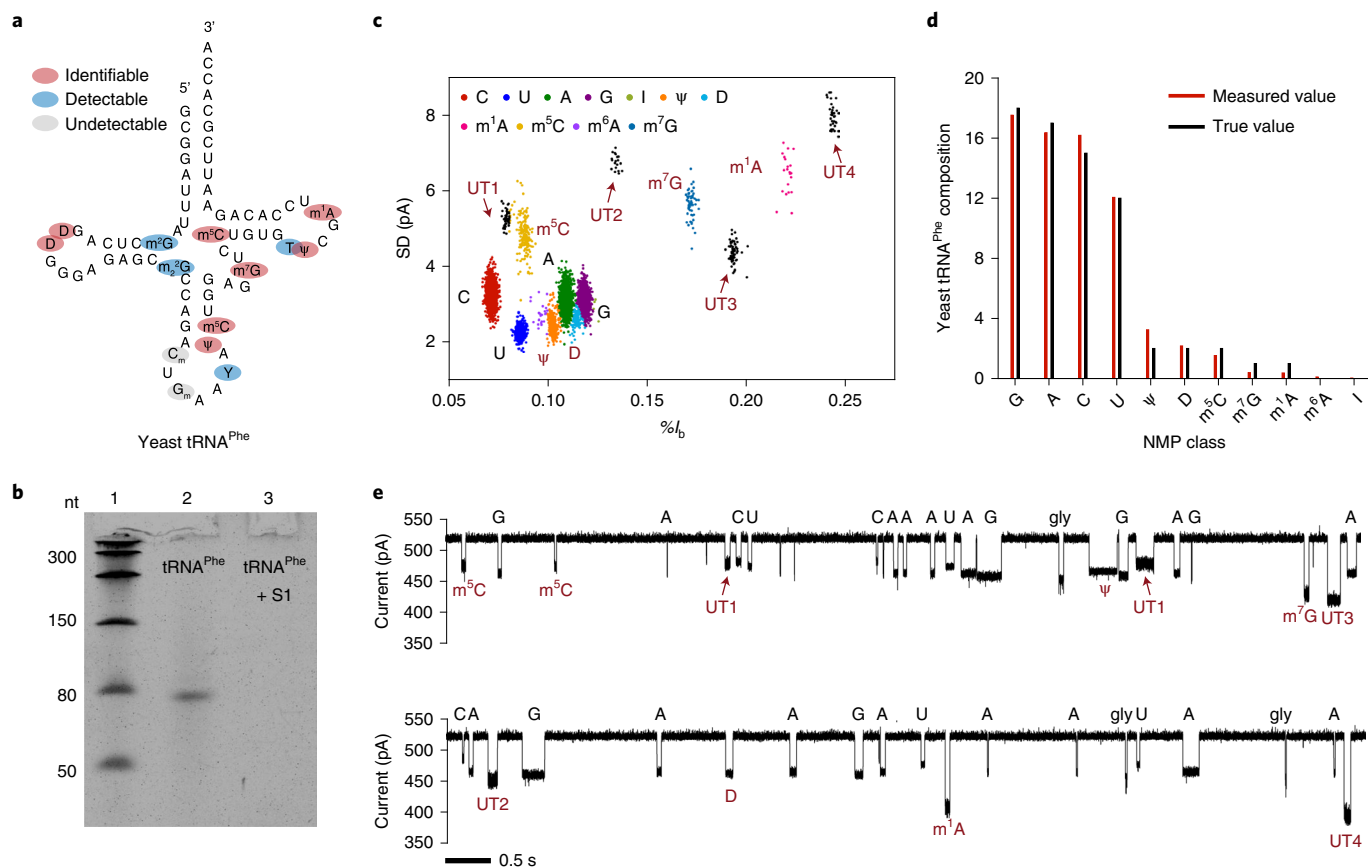


Fig. 5 | Quantitative detection of epigenetic modifications of yeast tRNA^{Phe}. **a**, The sequence and the modifications of a yeast tRNA^{Phe}. Positions of modifications, including m²G, D, m²G, C_{mr}, G_{mr}, ψ, T, m⁵C, m⁷G and m¹A, are shaded. These modifications can be divided into three categories. D, ψ, m⁵C, m⁷G and m¹A (shaded with red circles), which have known event features, are identifiable. m²G, m²G, T and Y (shaded with blue circles) are in principle detectable by MspA-PBA but are not identifiable. Due to a lack of *cis*-diol, C_m and G_m (shaded with grey circles) are in principle not detectable by MspA-PBA. **b**, Gel electrophoresis results. Lane 1: Low range ssRNA ladder (New England Biolabs). Lane 2: yeast tRNA^{Phe}; Lane 3: yeast tRNA^{Phe} treated with S1 nuclease. The gel result shows that the yeast tRNA^{Phe} were completely digested by S1 nuclease treatment. Operations of yeast tRNA^{Phe} digestion are detailed in Methods. **c**, The scatter plot of %I_b versus SD for tRNA^{Phe} digestion products. Events demonstrated were acquired from a 240 min continuous recording. The NMP identity is predicted by the Linear SVM model. Nine main populations corresponding, respectively, to events of C, U, A, G, m⁵C, ψ, D, m⁷G and m¹A were identified. Four populations of events, which do not belong to any previously identified event type, were also detected by the unsupervised learning method DBSCAN (Supplementary Fig. 37). UT stands for unidentified events. **d**, Comparison of the yeast tRNA^{Phe} composition between that derived from measurements and the true value. The measured values were determined and calibrated according to Methods. **e**, Representative traces acquired during nanopore sensing of the tRNA^{Phe} digestion products. Characteristic events of canonical and epigenetic NMPs are identified by machine learning and are marked with the corresponding labels. All measurements were carried out as described in Methods in a 1.5 M KCl buffer (1.5 M KCl, 10 mM MOPS, pH 7.0). A transmembrane potential of +200 mV was continually applied. Yeast tRNA^{Phe} digestion product was added to *cis* with a final concentration of 100 ng μl⁻¹.

of the S1 nuclease, were further removed from the dataset by machine learning (Supplementary Fig. 35). To cope with unknown epigenetic modifications in yeast tRNA^{Phe}, we combined supervised and unsupervised learning algorithms to identify the remaining events of NMPs. Here, One-Class SVM was employed to recognize events that do not belong to any previously trained event types. These events are considered as outliers. On the contrary, events that match the previously trained event types are considered as inliers (Supplementary Fig. 36), which are further identified by the trained Linear SVM model. The outlier events were, however, analysed with a density-based spatial clustering of applications with noise model (DBSCAN) to detect events appearing as clusters (Supplementary Fig. 37). The non-clustered events, which randomly distributed in the scatter plot, are considered as background events and are removed from the dataset without further analysis.

The result of the modification profile of yeast tRNA^{Phe} is shown in Fig. 5c. D, ψ, m⁵C, m⁷G and m¹A were successfully detected, consistent with the previous training results and literature⁴⁶.

Few m⁶A events were observed, which may be from background events that coincidentally share a similar event feature of m⁶A or other types of RNA mixed in the sample. Four new clusters of events, which demonstrated event features different from all NMP types that were previously applied for training, were also observed. These new clusters of events are likely from the m²G, m²G, T, Y or other unknown modifications in yeast tRNA^{Phe}. Quantitative analysis shows that the relative NMP composition in yeast tRNA^{Phe} is 17.53 GMP, 16.36 AMP, 16.19 CMP, 12.06 UMP, 3.24 ψ, 2.17 D, 1.53 m⁵C, 0.40 m⁷G, 0.37 m¹A, 0.11 m⁶A and 0.04 I, generally in accordance with the calculated true values (Fig. 5d)⁴⁶. A total of three independent trials was also performed (Supplementary Fig. 38) and the same conclusion was drawn, confirming the repeatability of this technique. Representative traces containing events of the yeast tRNA^{Phe} digestion products are also presented in Fig. 5e. With the above results, the capacity of MspA-PBA to measure NMPs and their epigenetic modifications from natural RNAs have been well approved.

Conclusions

A hetero-octameric MspA containing a sole PBA is reported to sense NMPs. Eleven types of NMPs are fully distinguished, outperforming those demonstrated by α -HL^{32,33} or solid-state nanopore^{47,48}. A machine learning algorithm was built, reporting a 0.996 accuracy. This work reports the largest number of NMP types that can be fully distinguished using nanopore. The only limitation is that the current sensing strategy fails to detect ribose-modified NMPs, such as C_m and G_m (ref. ¹). Compared with mass spectrometry, our method offers a higher resolution, especially in distinguishing RNA positional isomers (Supplementary Fig. 39). It is thus more suitable for RNA modification detections from mixed and native samples, without a need to couple with any chromatographic separation technology and complex data interpretation. This sensing strategy was also applied to identification of enzymatically cleaved NMPs from native RNA samples, suggesting the feasibility of *exo*-sequencing using enzyme-conjugated MspA-PBA. Although not demonstrated, this strategy is in principle suitable for sensing nucleoside diphosphates, nucleoside triphosphates, other nucleotide modifications, nucleotide sugars and nucleoside drugs, as long as the *cis-diol* of the ribose is still retained.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41565-022-01169-2>.

Received: 31 January 2022; Accepted: 1 June 2022;

Published online: 18 July 2022

References

- Boccalletto, P. et al. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* **46**, D303–D307 (2018).
- Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA modifications in gene expression regulation. *Cell* **169**, 1187–1200 (2017).
- Hausmann, I. U. et al. m6A potentiates Sxl alternative pre-mRNA splicing for robust *Drosophila* sex determination. *Nature* **540**, 301–304 (2016).
- Yang, X. et al. 5-methylcytosine promotes mRNA export — NSUN2 as the methyltransferase and ALYREF as an m5C reader. *Cell Res.* **27**, 606–625 (2017).
- Helm, M. Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res.* **34**, 721–733 (2006).
- Liu, J. et al. N 6-methyladenosine of chromosome-associated regulatory RNA regulates chromatin state and transcription. *Science* **367**, 580–586 (2020).
- Barbieri, I. & Kouzarides, T. Role of RNA modifications in cancer. *Nat. Rev. Cancer* **20**, 303–322 (2020).
- Bednářová, A. et al. Lost in translation: Defects in transfer RNA modifications and neurological disorders. *Front. Mol. Neurosci.* **10**, 135 (2017).
- Jonkhout, N. et al. The RNA modification landscape in human disease. *RNA* **23**, 1754–1769 (2017).
- Yu, Q. et al. RNA demethylation increases the yield and biomass of rice and potato plants in field trials. *Nat. Biotechnol.* **39**, 1581–1588 (2021).
- Ontiveros, R. J., Stoute, J. & Liu, K. F. The chemical diversity of RNA modifications. *Biochem. J.* **476**, 1227–1245 (2019).
- Keith, G. Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. *Biochimie* **77**, 142–144 (1995).
- Xu, J., Gu, A. Y., Thumati, N. R. & Wong, J. M. Y. Quantification of pseudouridine levels in cellular RNA pools with a modified HPLC–UV assay. *Genes (Basel)* **8**, 219 (2017).
- Wetzel, C. & Limbach, P. A. Mass spectrometry of modified RNAs: recent developments. *Analyst* **141**, 16–23 (2016).
- Li, X., Xiong, X. & Yi, C. Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat. Methods* **14**, 23–31 (2017).
- Linder, B. et al. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* **12**, 767–772 (2015).
- Schaefer, M., Pollex, T., Hanna, K. & Lyko, F. RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic Acids Res.* **37**, e12–e12 (2009).
- Carlile, T. M. et al. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**, 143–146 (2014).
- Hu, L. et al. m6A RNA modifications are measured at single-base resolution across the mammalian transcriptome. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-022-01243-z> (2022).
- Edelheit, S., Schwartz, S., Mumbach, M. R., Wurtzel, O. & Sorek, R. Transcriptome-wide mapping of 5-methylcytosine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genet.* **9**, e1003602 (2013).
- Dominissini, D. et al. The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530**, 441–446 (2016).
- Enroth, C. et al. Detection of internal N7-methylguanosine (m7G) RNA modifications by mutational profiling sequencing. *Nucleic Acids Res.* **47**, e126–e126 (2019).
- Delatte, B. et al. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science* **351**, 282–285 (2016).
- Arango, D. et al. Acetylation of cytidine in mRNA promotes translation efficiency. *Cell* **175**, 1872–1886 (2018).
- Okada, S., Ueda, H., Noda, Y. & Suzuki, T. Transcriptome-wide identification of A-to-I RNA editing sites using ICE-seq. *Methods* **156**, 66–78 (2019).
- Zhao, L. et al. Analysis of transcriptome and epitranscriptome in plants using pacbio Iso-seq and nanopore-based direct RNA sequencing. *Front. Genet.* **10**, 253 (2019).
- Vilfan, I. D. et al. Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. *J. Nanobiotechnol.* **11**, 8 (2013).
- Smith, A. M., Jain, M., Mulroney, L., Garalde, D. R. & Akeson, M. Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. *PLoS ONE* **14**, e0216709 (2019).
- Fleming, A. M., Mathewson, N. J., Howpay Manage, S. A. & Burrows, C. J. Nanopore dwell time analysis permits sequencing and conformational assignment of pseudouridine in SARS-CoV-2. *ACS Cent. Sci.* **7**, 1707–1717 (2021).
- Goodwin, S. et al. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res.* **25**, 1750–1756 (2015).
- Begik, O. et al. Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nat. Biotechnol.* **39**, 1278–1291 (2021).
- Ayub, M., Hardwick, S. W., Luisi, B. F. & Bayley, H. Nanopore-based identification of individual nucleotides for direct RNA sequencing. *Nano Lett.* **13**, 6144–6150 (2013).
- Clarke, J. et al. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **4**, 265–270 (2009).
- Song, L. et al. Structure of staphylococcal α -hemolysin, a heptameric transmembrane pore. *Science* **274**, 1859–1865 (1996).
- Faller, M., Niederweis, M. & Schulz, G. E. The structure of a mycobacterial outer-membrane channel. *Science* **303**, 1189–1192 (2004).
- Manrao, E. A. et al. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* **30**, 349–353 (2012).
- Cao, J. et al. Giant single molecule chemistry events observed from a tetrachloroaurate(III) embedded Mycobacterium smegmatis porin A nanopore. *Nat. Commun.* **10**, 5668 (2019).
- Wang, Y. et al. Structural-profiling of low molecular weight RNAs by nanopore trapping/translocation using Mycobacterium smegmatis porin A. *Nat. Commun.* **12**, 3368 (2021).
- Springsteen, G. & Wang, B. A detailed examination of boronic acid–diol complexation. *Tetrahedron* **58**, 5291–5300 (2002).
- Ramsay, W. J. & Bayley, H. Single-molecule determination of the isomers of d-Glucose and d-Fructose that bind to boronic acids. *Angew. Chem.* **130**, 2891–2895 (2018).
- Jia, W. et al. Programmable nano-reactors for stochastic sensing. *Nat. Commun.* **12**, 5811 (2021).
- Choi, L.-S. & Bayley, H. S-Nitrosothiol chemistry at the single-molecule level. *Angew. Chem. Int. Ed.* **51**, 7972–7976 (2012).
- Yurkevich, A. M. et al. The reaction of phenylboronic acid with nucleosides and mononucleotides. *Tetrahedron* **25**, 477–484 (1969).
- Konno, M. et al. Distinct methylation levels of mature microRNAs in gastrointestinal cancers. *Nat. Commun.* **10**, 3888 (2019).
- Hori, H. Methylated nucleosides in tRNA and tRNA methyltransferases. *Front. Genet.* **5**, 144 (2014).
- Hingerty, B., Brown, R. & Jack, A. Further refinement of the structure of yeast tRNA^{Phe}. *J. Mol. Biol.* **124**, 523–534 (1978).
- Feng, J. et al. Identification of single nucleotides in MoS₂ nanopores. *Nat. Nanotechnol.* **10**, 1070–1076 (2015).
- Jeong, K.-B. et al. Alpha-Hederin nanopore for single nucleotide discrimination. *ACS Nano* **13**, 1719–1727 (2019).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2022

Methods

Preparation of homo-octameric MspAs. The genes coding for monomeric M2 MspA-D16H6 (D90N/D91N/D93N/D118R/D134R/E139K) and N90C MspA-H6 (D90C/D91N/D93N/D118R/D134R/E139K) were separately synthesized and simultaneously inserted into a pET 30a(+) plasmid (GenScript). A hexa-histidine tag (H6), which assists purification by nickel affinity chromatography, was added to the C terminus of both genes. A 16 aspartate tag (D16) was added to the end of the M2 MspA-D16H6 gene to enhance discrimination during gel electrophoresis between octameric M2 MspA-D16H6 and N90C MspA-H6.

The preparation of homo-octameric M2 MspA-D16H6 and N90C MspA-H6 was performed as previously reported⁴⁹. Experimentally, 100 ng of either recombinant plasmid was added to 100 μ l of *Escherichia coli* BL21 (DE3) pLysS competent cells (Sangon Biotech) and incubated on ice for 30 min. After heat shock transformation performed at 42 °C for 90 s, the mixture was cultured on ice for another 3 min. Then the mixture was added to 800 μ l LB broth and shaken at 37 °C and 175 r.p.m. for 50 min. Subsequently, the mixture was spread onto a LB agar plate containing kanamycin (30 μ g ml⁻¹) and chloramphenicol (34 μ g ml⁻¹) and cultured for 18 h. A single colony was inoculated and added to 100 ml LB broth containing kanamycin (30 μ g ml⁻¹) and chloramphenicol (34 μ g ml⁻¹) in a 250 ml flask. The mixture was shaken at 37 °C and 175 r.p.m. until the optical density at 600 nm (OD₆₀₀) reached 0.7. Isopropyl β -D-1-thiogalactopyranoside (IPTG) was then added to a final concentration of 0.5 mM to induce protein expression. The medium was shaken at 16 °C and 175 r.p.m. for a further 16 h. Finally, the medium was centrifuged at 4,000 r.p.m. and 4 °C for 20 min to collect the cell pellet.

The collected bacterial pellet was resuspended in 40 ml of a lysis buffer (100 mM Na₂HPO₄/NaH₂PO₄, 0.1 mM ethylenediaminetetraacetic acid (EDTA), 150 mM NaCl, 0.5% (v/v) Genapol X-80, pH 6.5) and incubated at 60 °C for 10 min. Afterwards, the suspension was ice-incubated for 10 min. The suspension was centrifuged at 13,000 r.p.m. for 40 min at 4 °C. The supernatant was collected and filtered with a 0.2 μ m syringe filter (Nalgene). The filtered solution was then loaded to a HisTrap HP nickel ion affinity column (GE Healthcare). The column was first eluted with buffer A (0.5 M NaCl, 20 mM HEPES, 5 mM imidazole, 0.5% (v/v) Genapol X-80, pH 8.0) until the UV absorbance stabilized. It was then eluted using a linear gradient of buffer B (0.5 M NaCl, 20 mM HEPES, 500 mM imidazole, 0.5% (v/v) Genapol X-80, pH 8.0) and buffer A over six column volumes within 30 min. Tris(2-carboxyethyl) phosphine (TCEP) was added to both buffer A and buffer B with a final concentration of 2 mM to prevent the formation of disulfide bonds between cysteine residues when purifying homo-octameric N90C MspA-H6 (ref. 50). Finally, the eluted fractions were separately collected and characterized by gel electrophoresis (4–20% gradient sodium dodecyl sulfate (SDS)–polyacrylamide gel). The fractions containing the desired product were stored at –80 °C for subsequent use.

Preparation of (N90C)₁(M2)₇. For simplicity, the hetero-octameric MspA, which is composed of one fraction of N90C MspA-H6 and seven fractions of M2 MspA-D16H6, is referred to as (N90C)₁(M2)₇. To prepare for (N90C)₁(M2)₇, the genes coding for N90C MspA-H6 and M2 MspA-D16H6 were simultaneously placed in a co-expression vector pETDuet-1 (Supplementary Fig. 1). Specifically, the gene coding for N90C MspA-H6 was inserted between the restriction sites of NcoI and HindIII. The gene coding for M2 MspA-D16H6 was inserted between the restriction sites of NdeI and BlnI. A hexa-histidine tag (H6) was added to the C terminus of both genes to assist purification by nickel affinity chromatography. A 16 aspartate tag (D16) was added to the end of the M2 MspA-D16H6 gene to enhance the discrimination between hetero-octameric MspAs during gel electrophoresis.

Experimentally, 100 ng recombinant plasmid was transformed into 100 μ l *E. coli* BL21 (DE3) pLysS competent cells (Sangon Biotech) and cultured on ice for 30 min. After heat shock transformation performed at 42 °C for 90 s, the mixture was cultured on ice for another 3 min. Then the mixture was added to 800 μ l LB broth and cultured at 37 °C and 175 r.p.m. for 50 min. Subsequently, the mixture was spread onto a LB agar plate containing ampicillin (50 μ g ml⁻¹) and chloramphenicol (34 μ g ml⁻¹) and cultured for 18 h. A single colony was inoculated and added to LB broth containing ampicillin (50 μ g ml⁻¹) and chloramphenicol (34 μ g ml⁻¹). The mixture was shaken at 37 °C and 175 r.p.m. until OD₆₀₀ reached 0.7. The medium was then transferred to 1 l LB broth containing ampicillin (50 μ g ml⁻¹) and chloramphenicol (34 μ g ml⁻¹). The mixture was shaken at 37 °C and 175 r.p.m. until OD₆₀₀ reached 0.6. To induce protein expression, IPTG was then added to a final concentration of 0.1 mM. The medium was shaken at 16 °C and 175 r.p.m. for another 24 h. Finally, the medium was centrifuged at 4,000 r.p.m. for 20 min at 4 °C to collect the bacterial pellet.

The collected bacterial pellet was resuspended in 160 ml of lysis buffer (100 mM Na₂HPO₄/NaH₂PO₄, 0.1 mM EDTA, 150 mM NaCl, 0.5% (v/v) Genapol X-80, pH 6.5) and incubated at 60 °C for 50 min. After ice-incubation for 30 min, the suspension was centrifuged at 13,000 r.p.m. for 40 min at 4 °C. The supernatant was collected and filtered with a 0.2 μ m syringe filter (Nalgene). It was then loaded to a HisTrap HP nickel ion affinity column (GE Healthcare). The column was first eluted with buffer A (0.5 M NaCl, 20 mM HEPES, 5 mM imidazole, 2 mM TCEP, 0.5% (v/v) Genapol X-80, pH 8.0) until the UV absorbance reached a stable level. It was then eluted using a linear gradient of buffer B (0.5 M NaCl, 20 mM HEPES, 500 mM imidazole, 2 mM TCEP, 0.5% (v/v) Genapol X-80, pH 8.0) and buffer A over 12 column volumes within 60 min. The elution fractions were separately collected and characterized by gel electrophoresis on a 4–20% gradient

SDS–polyacrylamide gel (Supplementary Fig. 2). The fractions corresponding to all hetero-octameric MspAs were collected for further purifications.

Further separation of hetero-octameric MspA was performed on a 10% SDS–polyacrylamide gel (Supplementary Fig. 3). Gel electrophoresis was continually run for 16 h with a +160 V applied potential. The gel was then stained with coomassie brilliant blue (1.25 g coomassie brilliant blue R250, 225 ml methanol, 50 ml glacial acetic acid, 225 ml ultrapure water) for 4 h. Subsequently, it was immersed with the de-staining buffer (400 ml methanol, 100 ml glacial acetic acid, replenished with ultrapure water to 1 l) until the protein bands were clearly visible. The protein band which corresponds to (N90C)₁(M2)₇ was excised from the gel and immersed with an extraction solution (150 mM NaCl, 15 mM Tris-HCl, pH 7.5, 0.2% DDM, 0.5% Genapol X-80, 5 mM TCEP, 10 mM EDTA) for 12 h. The mixture was collected and stored at –80 °C for subsequent use.

Preparation of MspA modified with a PBA. To modify (N90C)₁(M2)₇ with a phenylboronic acid, 1 μ l prepared (N90C)₁(M2)₇, 0.2 μ l MPBA (1 M, dissolved in dimethyl sulfoxide) and 8.8 μ l 1.5 M KCl buffer (1.5 M KCl, 10 mM MOPS, pH 7.0) were mixed and incubated for 10 min. For simplicity, the PBA-modified MspA is referred to as MspA-PBA throughout the paper, if not otherwise stated.

Nanopore measurements. Nanopore measurements were performed similarly to that described previously⁵⁰. To avoid interference from the measurement environment, the custom-made measurement device was fixed in a homemade Faraday cage mounted on an optical table (Jiangxi Liansheng technology Co., Ltd). The liquid chamber of the measurement device was separated by a Teflon film containing a 100 μ m diameter orifice. Before each use, the orifice was first treated with 0.5% (v/v) hexadecane in pentane. Both chambers were then filled with 500 μ l KCl buffer (1.5 M KCl, 10 mM MOPS, pH 7.0). A pair of Ag/AgCl electrodes, which were electrically extended from a patch-clamp amplifier, was immersed in both chambers, in contact with the buffers. Conventionally, the chamber that is electrically grounded is defined as *cis* and its opposing chamber is defined as *trans*. By adding a drop of 5 mg ml⁻¹ DPhPC in pentane to each chamber and pipetting the liquid in either chamber up and down several times, the lipid bilayer was spontaneously formed. Then, octameric (N90C)₁(M2)₇ or MspA-PBA was added to *cis* to trigger spontaneous pore insertion into the lipid bilayer. To avoid further insertions, the buffer in *cis* was exchanged with fresh buffer, upon a single nanopore insertion.

All single channel recordings were performed with an Axonpatch 200B patch-clamp amplifier coupled with a Digidata 1550B digitizer. The sampling rate is 25 kHz and the acquired trace is further digitally low-pass filtered with a corner frequency of 1 kHz. Unless otherwise stated, a +200 mV voltage was continually applied during all measurements. All analytes were added to the *cis* chamber to a desired concentration.

Data analysis. Raw Axon abf files were imported into MATLAB using the 'abfload' function downloaded from <https://www.mathworks.com/matlabcentral/fileexchange/6190-abfload>. The characteristic parameters of each event including %I_b, SD, t_{off} and t_{on} were extracted with a custom MATLAB program. Events with a t_{off} < 10 ms were ignored. Subsequent analyses including histogram plots, scatter plots, violin plots and curve fittings were performed in Origin v9.1 (Origin Lab).

Machine learning was performed by MATLAB. Five-hundred events of each analyte type were collected to form a dataset. The label for each event was assigned with the known identity of the analyte. The dataset was then split into a training set (80%) and a testing set (20%) for model training and model testing. %I_b and SD of events were employed as event features. Model training was performed using the Classification Learner toolbox of MATLAB. Mainstream classifiers including Decision Trees, Discriminant Analysis, Naïve Bayes, SVM, K Nearest Neighbour, Ensemble and Neural Network were estimated with default settings. According to results of tenfold cross validation accuracy and the testing accuracy, the Linear SVM model demonstrated the best performance. A confusion matrix and decision boundary were generated based on results of the Linear SVM model. The trained model was then applied for predictions of unlabelled data.

One-Class SVM was performed by MATLAB. Five-hundred events of each analyte type were collected to form a dataset. %I_b and SD of events were employed as event features. The OutlierFraction was set to 0.0005. Density-based spatial clustering of applications with noise cluster analysis was performed with Python. Parameters of Epsilon was set to 0.12 and min_samples was set to 18.

MicroRNA digestion. S1 nuclease (Takara) was applied to enzymatically digest RNA into nucleoside monophosphates (NMPs). Before the digestion, S1 nuclease was pretreated by ultrafiltration (Amicon, Ultra-0.5 ml, Ultracel-10 K) to remove glycerol. After ultrafiltration, the remaining solution in the filter device which contained S1 nuclease was collected. Subsequently, the reaction was performed by mixing 150 μ g microRNA, 21 μ l pretreated S1 nuclease solution (180 U μ l⁻¹), 6 μ l 10X S1 nuclease buffer (300 mM CH₃COONa, 2,800 mM NaCl, 10 mM ZnSO₄, pH 4.6) and ultrapure water to a final volume of 60 μ l. The reaction was kept at 23 °C for 4 h. To separate digested products, the mixture was then added to a centrifugal filter with a 10 kDa molecular weight cut off (MWCO) and centrifuged at 8,000 r.p.m. for 60 min at 4 °C. The filtrate was collected and stored at 4 °C for subsequent uses. All tips and tubes used were RNase-free.

Yeast tRNA^{Phe} digestion. S1 nuclease (Takara) was applied to enzymatically digest RNA into nucleoside monophosphates (NMPs). Before the digestion, S1 nuclease was pretreated by four turns of ultrafiltration (Amicon, Ultra-0.5 ml, Ultracel-10 K) to remove glycerol. During each centrifugation operation, the S1 nuclease solution was added to the centrifugal filter with a 10 kDa MWCO and centrifuged at 8,000 r.p.m. for 60 min at 4 °C. After ultrafiltration, the remaining solution in the filter device which contained the S1 nuclease was collected. Subsequently, the reaction was performed by mixing 50 µg yeast RNA^{Phe}, 28 µl pretreated S1 nuclease solution (180 U µl⁻¹), 8 µl 10X S1 nuclease buffer (300 mM CH₃COONa, 2,800 mM NaCl 10 mM ZnSO₄, pH 4.6) and ultrapure water to a final volume of 80 µl. The reaction was kept at 23 °C for 15 h. To separate the digested products, the mixture was then added to a centrifugal filter with a 10 kDa MWCO and centrifuged at 8,000 r.p.m. for 60 min at 4 °C. The filtrate was collected and vacuum dried for 6 h. The powder was stored at 4 °C for subsequent uses. All tips and tubes used were RNase-free.

RNA composition quantification. During nanopore sensing of RNA digestion products, the digested NMP concentrations (C_i) were evaluated according to the following equation:

$$C_i = E_i / (\delta_i \times t)$$

Here, the annotation i (from 1 to 11) stands for parameters relevant to CMP, UMP, AMP, GMP, m^cC, m^aA, ψ, I, D, m^gG and m¹A, respectively. Here, E_i is the number of corresponding NMP binding events detected during a continuous sensing of RNA digestion products. An example of E_i is shown in Supplementary Fig. 33b,d. δ_i is the calibration coefficient, which is defined as the number of NMP binding events occurring per unit concentration (µM) per min. The values of δ_i were acquired during measurements with 300 µM corresponding NMP at +200 mV. δ_i are also summarized in Supplementary Table 10. t is the recording time, of 60 min.

The nucleotide compositions of RNA were derived according to the following equation:

$$N_i = L \frac{C_i}{\sum_{i=1}^{11} C_i}$$

Here, L is the length of the RNA. For hsa-miR-17, $L = 23$. For hsa-miR-21, $L = 22$. For yeast tRNA^{Phe}, $L = 70$.

Data availability

The datasets generated and/or analysed during the current study are available within the source data provided with this paper. All data presented in this work can be requested from the corresponding author upon reasonable request.

Code availability

The custom machine learning code is shared on GitHub as 'NMP classifier' at <https://github.com/sonic220/NMP-Classifier>.

References

- Wang, Y. et al. Osmosis-driven motion-type modulation of biological nanopores for parallel optical nucleic acid sensing. *ACS Appl. Mater. Interfaces* **10**, 7788–7797 (2018).
- Wang, S. et al. Single molecule observation of hard–soft–acid–base (HSAB) interaction in engineered *Mycobacterium smegmatis* porin A (MspA) nanopores. *Chem. Sci.* **11**, 879–887 (2020).

Acknowledgements

We acknowledge H. Bayley (University of Oxford) for valuable suggestions concerning preparation of the manuscript. We acknowledge Z. Guo, S. Zhu, C. Zhu, J. Li, R. Xie, Y. Guo, X. Chen and J. Xu in Nanjing University for inspiring discussions. This project was funded by the National Natural Science Foundation of China (grant Nos. 31972917, 91753108 and 21675083, to S.H.), and supported by the Fundamental Research Funds for the Central Universities (grant Nos. 020514380257 and 020514380261, to S.H.), programmes for high-level entrepreneurial and innovative talents introduction of Jiangsu Province (individual and group programme, to S.H.), the Natural Science Foundation of Jiangsu Province (grant No. BK20200009, to S.H.), the Excellent Research Programme of Nanjing University (grant No. ZYJH004, to S.H.), the Shanghai Municipal Science and Technology Major Project (S.H.), the State Key Laboratory of Analytical Chemistry for Life Science (grant No. 5431ZZXM2204, to S.H.), the Technology innovation fund programme of Nanjing University (S.H.) and the China Postdoctoral Science Foundation (grant No. 2021M691508, to Y.Q.W.).

Author contributions

S.H., Y.Q.W., S.Y.Z. and J.W.D. conceived the project. Y.Q.W., S.Y.Z., P.P.F., K.F.W. and X.Y.D. prepared the MspA nanopores. Y.Q.W., S.Y.Z., P.P.F., L.Y.W., X.Y.L., J.L.C., Z.Y.C. and C.Z.H. performed the measurements. Y.Q.W. and Y.L. designed the machine learning algorithms. Y.Q.W. and J.Y.Z. performed RNA extraction. Y.Q.W. and J.H. performed the mass spectroscopy measurement. P.K.Z. set up the instruments. S.H. and Y.Q.W. wrote the paper. S.H. and H.Y.C. supervised the project.

Competing interests

S.H., S.Y.Z., Y.Q.W., K.F.W. and Y.L. have filed patents describing the preparation of heterogeneous MspA and its applications thereof. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41565-022-01169-2>.

Correspondence and requests for materials should be addressed to Shuo Huang.

Peer review information *Nature Nanotechnology* thanks Sukanya Punthambaker and Manisha Gupta for their contribution to the peer review of this work.

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