nature portfolio

Peer Review File

An engineered nanopore identifies saccharides, amino acids, peptides, and ribonucleotides

Corresponding Author: Professor Shuo Huang

Parts of this Peer Review File have been redacted as indicated as we could not obtain permission to publish the reports from one of the peer reviewers.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In this NBT manuscript, Yao and colleagues developed a novel Mycobacterium smegmatis porin (MspA)-based nanopore sensor capable of distinguishing 21 amino acids, four canonical nucleobases, four monosaccharides, and beyond, including post-translational modifications (PTMs). These PTMs include modifications of residues and peptides. The overall design of this sensor involves conjugating maleimido C2-formylphenylboronic acid (maleimido-C2-FPBA) to an MspA nanopore. The authors strategically positioned the conjugation site within the constriction of the MspA nanopore. By combining nanopore technology with advanced machine learning algorithms, they achieved an overall accuracy of 98.7%.

Additionally, the introduction of few-shot learning enhanced the detection of unknown or “out-of-database” events and their corresponding analytes. The nanopore sensor was also tested in complex biofluids, with experiments conducted in yeast cell extract. While engineered heteromeric MspA nanopores and the use of deep learning in single-molecule nanopore sensing are already well established, the “unification” of simultaneous sensing of multiple analytes appears to be novel, to the best of this reviewer's knowledge. Overall, the data are well organized and presented. However, the lack of clarity regarding some technical aspects of this work diminishes this reviewer's enthusiasm.

Major concerns:

- The authors are definitely aware of many recent studies related to this work by other research groups. However, there is no critical discussion comparing and contrasting with other nanopore detection methods, some of which are not even cited (e.g., Ouldali and coworkers, Nature Biotechnol, 38(2), 176-181, 2020). What specifically is improved compared to prior work?

- Additionally, there is no discussion of the resolution limit (such as the detection threshold) of these building blocks. Specifically, the reader would probably want to understand what potential challenges could prevent the immediate application of this approach in real-world biotechnological settings. This point is significant, as NBT is currently reviewing the paper.

- There is no indication of the filter used to produce the electrical traces in each figure. The methods mention that events shorter than 2 ms were removed, but was additional filtering of the data conducted to produce these plots, as they are relatively clean traces? For example, in Figure 6, I expected noisier traces, given it was conducted in yeast extract. If there was additional filtering, then these traces would be more obvious to me. I imagine that other components of the extract could clog the pore. Was there additional data processing or experimental handling required to maintain the detection resolution? If so, those technical details and discussions need to be added.

- It would be helpful to include an explanation of how the parameters SD and ΔI were chosen as the key identifying elements. The authors list all the parameters analyzed and provide them in supplementary tables, but a more detailed explanation of whether other parameters could also be used to distinguish these analytes would be beneficial.

- The overall approach should definitely be benchmarked against other competing technologies. For example, in the glycosylated peptide experiments, additional enzymatic degradation of a non-glycosylated peptide would be necessary. Alternatively, the authors could explain or provide additional data to demonstrate their confidence that the three unknown events are not artifacts of enzymatic fragmentation.

Minor comments.
- The first sentence of the abstract is the same as the first sentence of the main text.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Biotechnology initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
The manuscript by Yao et al focuses on analysis of various small-molecule agents using a clever pore modification. The authors have modified an MspA nanopore with a formyl-phenylboronic acid moiety to form a pore they call "MspA-FPBA". This engineered pore, which contains both a boronic acid moiety and an aldehyde residue, was then used to detect all of the 21 canonical amino acids, as well as the RNA mononucleoside monophosphates (NMPs), and some sugars. The basis of these detections are that: 1) α-amino acids interact with aldehydes by forming a Schiff Base with the α-amino groups, which results in a long dwell time of interaction (~1 s). 2) NMPs interact via reversible boronic acid/diol interactions, and 3) sugars interact the same way (through diol/boronic acid interactions). Altogether, almost 40 molecules are detected and distinguished using this scheme (among these molecules are short peptides). For such small molecules (MW < 1 kD), 1 second long interaction times with the pores is extremely long - in fact, one would expect to find unobservable sub-μs long dwell times. The reversible chemical interaction, as well as the careful positioning of the anchoring groups in the pore constriction, are responsible for the observations of unique ion current signatures.

Full disclaimer: this paper is a very impressive demonstration of the usability of nanopores for discrimination among various small molecule species. The writing is impeccable, and the results are clearly presented. I have no comments on the rigor and writing.

Having said that, however, I must admit that I have a hard time with its positioning in the realm of usability and novelty for Nature Biotechnology. One major issue here is that all of these molecules are detected at μM to mM concentrations, and mass spectrometers can equally do a good job at detecting and quantifying these molecules and more in a mixture. Second, the authors have already demonstrated the detection of amino acids as well as sugars and NTPs in previous papers. So, this paper feels a bit like an incremental achievement to the previous works by the authors (all of which I genuinely admire).

Therefore, while I have no technical comments about the data or conclusions made in the manuscript, I believe that this work is more suitable when adjoined to a more tangible application that would highlight the importance of detecting all of these species (or some of them) in a complex mixture. The yeast extract data in figure 6 is interesting, but falls short of quantification and exhaustive analysis that would highlight anything particularly useful to the bioanalytical field.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
The manuscript presents anintegration of machine learning with a cleverly modified single-molecule nanopore sensing platform for molecules. It presents a strategy to detect single biomolecule types (amino acids, post-translationally modified amino acid, nucleotides, epigenetically modified nucleotides, and glycans) as well as the same molecules in mixtures. The apparent strategy would be, in the future, to have a cleaving molecule such as a protease/nuclease/glycoside hydrolase at the entrance of the pore, and use that to sequentially degrade respective molecules to be able to sequence each of these molecules. The current work takes a step in this direction. My co-reviewers and I have the following comments on the manuscript. Our main concern is with the validation of nanopore detected analytes in mixtures from both the nanopore experiments and the machine learning perspectives.

1. Overall, the manuscript is very well done and provides appropriate details. The use of the unified sensing adaptor maleimido-C2-FPBA adaptor on MspA is definitely a significant advance. The demonstrations of single species detection is

convincing.

2. While experiments with mixtures provide clear signals in almost all cases that are classifiable, I worry about the “ground truth” in such cases. How do we know which species is going through for sure. We have the nanopore signal which indicates a specific species but there is no independent validation of this in the experiments. A possible way would be to have specific mixtures (varying concentrations of species) and then through significant number of reads through the nanopores try to reproduce the concentration of species. This may or may not be feasible but some demonstration like this is needed. If the above is feasible, this may also be a way of determining the overall sequence of protease treated protein mixtures (for example) by combining experiments with ML training on single proteins solutions and multiple mixtures of proteins of increasing complexity.

The machine-learning methodology is generally appropriate in the context of this work and contributes to the strength of the study. Our specific comments are below:

1. The use of bagged trees ensemble achieves a near-99% accuracy under cross-validation, which is plausible given the distinct blockage currents for most of the molecules. The paper demonstrates an effective use of classical ML approaches combined with well-designed features to distinguish the signals from different molecules. However, our own reproduction of the model using the provided code files, suggests that a smaller subset of features {'Mean', 'SD', 'toff', 'LocalSD_Ratio'} is capable of achieving performance similar to the 9-feature model. Furthermore, fitting the provided data to Random Forest models shows a capability to achieve similar scores using pretty low n_estimators (3-5), and max_depth (10-20) parameters. Such near-maximal accuracy with such a lightweight model strongly suggests that the underlying feature space is highly separable and that several of the nine engineered features may be redundant. A brief analysis of the minimal feature subsets would significantly strengthen claim of the multi-functional nanopore producing a consistently informative and discriminative signals.

2. The code files provided only include the code and data for the AAs_NMPs_Sacs_classifier (used for Fig. 3 and Supplementary Table 6). The code and data used for the few-shot learning method described in Figure 5, as well as that which is used for identification of the PTM modified amino acids, epigenetic NMPs and peptides (and for Extended Data Figure 5) are missing from the source file. This makes it difficult to reproduce the study and verify the performance of the ML models. We recommend providing this to the reviewers and also to the readers eventually.

3. Lines 281-283: The authors note that the model used for the out-of-database classification was trained on the same dataset described in Fig. 3, which was generated exclusively from single-analyte experiments. While the few-shot generalization to new analytes is encouraging, extending a model trained solely on isolated analytes to scenarios involving multi-analyte mixtures requires additional justification. In multi-species experiments, event characteristics may be altered by competitive interactions, concentration imbalances, or subtle changes in translocation dynamics. Demonstrating that the event distributions from single-analyte experiments remain valid in these more complex conditions, would considerably strengthen the manuscript.

(Remarks on code availability)
See main review, some of the code was not made available and needs to be made available. We were able to run the code provided and provide some comments in the main review.

Reviewer #5

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Biotechnology initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #6

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Biotechnology initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Author Rebuttal letter:

Response Letter
Reviewer #1:

Review Comments:

In this NBT manuscript, Yao and colleagues developed a novel Mycobacterium smegmatis porin (MspA)-based nanopore sensor capable of distinguishing 21 amino acids, four canonical nucleobases, four monosaccharides, and beyond, including post-translational modifications (PTMs). These PTMs include modifications of residues and peptides. The overall design of this sensor involves conjugating maleimido C2-formylphenylboronic acid (maleimido-C2-FPBA) to an MspA nanopore. The authors strategically positioned the conjugation site within the constriction of the MspA nanopore. By combining nanopore technology with advanced machine learning algorithms, they achieved an overall accuracy of 98.7%.

Response:

We sincerely appreciate Reviewer 1's accurate characterization of our newly developed nanopore sensing system. We are particularly grateful that the reviewer recognized the strategic significance of this newly developped MspA-FPBA nanopore. This represents the first nanopore sensor capable of simultaneously detecting a broad spectrum of biomolecules, including amino acids, post-translationally modified amino acids, nucleoside monophosphates (NMPs), epigenetically modified NMPs, monosaccharides, and peptides, demonstrating remarkable versatility. We are deeply appreciative of the reviewer's recognition of the innovation and the extensive screening effort required to identify this universal reactive adapter. The resulting event features extracted from this sensor contain highly informative signal patterns that necessitate advanced machine learning algorithms for effective analysis. This computational approach enables automated, high-accuracy identification of diverse biomolecules, a capability we are especially gratified to see acknowledged by Reviewer 1.

Review Comments:

Additionally, the introduction of few-shot learning enhanced the detection of unknown or “out-of-database” events and their corresponding analytes. The nanopore sensor was also tested in complex biofluids, with experiments conducted in yeast cell extract. While engineered heteromeric MspA nanopores and the use of deep learning in single-molecule nanopore sensing are already well established, the “unification” of simultaneous sensing of multiple analytes appears to be novel, to the best of this reviewer's knowledge. Overall, the data are well organized and presented. However, the lack of clarity regarding some technical aspects of this work diminishes this reviewer's enthusiasm.
Response:

We sincerely appreciate your recognition of the innovative contributions presented in this work, including the integration of few-shot learning, the experimental validation in complex fluid environments, and the unification of simultaneous sensing capabilities. Your commendation regarding data presentation and the quality of the writing are particularly valued.

We have thoroughly addressed all your insightful questions through systematic methodological refinements and rigorous experimental validation. Below, we provide detailed point-by-point responses for your reference:

Review Comments:

The authors are definitely aware of many recent studies related to this work by other research groups. However, there is no critical discussion comparing and contrasting with other nanopore detection methods, some of which are not even cited (e.g., Ouldali and coworkers, Nature Biotechnol, 38(2), 176-181, 2020). What specifically is improved compared to prior work?

Response:

We highly acknowledge reviewer 1 for your valuable suggestions on the discussion of major prior studies relevant to this work. We sincerely apology for missing a thorough discussion on this. Strictly following your suggestion, in the revised manuscript, we have incorporated the relevant citations to ensure the comprehensiveness of our discussion, and have supplemented it with the technological improvements of this study.

Briefly, unlike conventional approaches, the newly developed MspA-FPBA sensor represents a substantial advancement by enabling multi-analyte detection all within a single assay system, simultaneously identifying amino acids, nucleoside monophosphates (NMPs), saccharides, and peptides. This multiplexed sensing platform is underpinned by the synthesis and comprehensive characterization of a novel small molecule refereed to as “maleimido-C2-formylphenylboronic acid”. The sensor's exceptional analytical versatility, combined with the rich event signatures produced and machine learning-based pattern recognition algorithms, enables “out-of-database” analyte classification. Rather than relying on exact event feature matching with perfect matching reference compounds, this innovative methodology identifies classification-relevant signal features that facilitate analyte-type grouping. This strategy holds particular promise for the discovery of novel macromolecular modifications, including but not limited to epigenetic marks, post-translational modifications (PTMs), and diverse glycosylation variants.

Finally, to fully address this comment, in the revised manuscript, we have added the discussion of relevant prior studies and core technical improvements made in this study. For the ease of your reference, these added texts are also pasted below.

“Nanopore identification of amino acids, NMPs, saccharides or peptides have been previously performed. For amino acid, phenanthroline ring-modified51 and cyclodextrin-embedded52 α-hemolysin (α-HL) identified amino acid enantiomers in the presence of Cu^{2+}. Single Ni^{2+}-modified41 and Cu^{2+}-modified53 MspA distinguished all 20 proteinogenic amino acids. For nucleotide, cyclodextrin-embedded α-HL detected four nucleoside diphosphates54. A single phenylboronic acid-modified MspA distinguished 11 nucleoside monophosphates40. For saccharide, single phenylboronic acid-modified α-HL distinguished three saccharides55, while similarly modified MspA differentiated major monosaccharides and
39.
disaccharides with different glycosidic linkages38, Mutant α-HL recognized saccharides bearing acetamido and carboxyl groups29. For peptide, Aerolysin56 and cucurbituril-assisted α-HL57 distinguished peptides differing by a single amino acid. Wild-type aerolysin58, the N. farcinica channel59, and the bacteriophage T7 connector60 differentiated arginine peptides of different lengths. Translocon EXP2 distinguished polylysine peptides with different molecular weights61. Fragaceatoxin C differentiated various angiotensin peptides62. Aerolysin distinguished vasopressin enantiomers63. α-HL identified multiple peptide post-translational modifications64. However, all above works were generally limited to a specific analyte class for each pore type. In contrast, MspA-FPBA enabled simultaneous identification of amino acids, NMPs, saccharides, and peptides, enabling their classification in the same platform. Assisted by machine learning, it could also perform analyte type classification without strictly relying on events generated by identical standard compounds.”

Review Comments:

Additionally, there is no discussion of the resolution limit (such as the detection threshold) of these building blocks. Specifically, the reader would probably want to understand what potential challenges could prevent the immediate application of this approach in real-world biotechnological settings. This point is significant, as NBT is currently reviewing the paper.

Response:

We sincerely appreciate reviewer 1 for the valuable comments. Based on our understanding of the reviewer’s comments, we infer that reviewer 1 requested an investigation of both the resolution limit and the detection threshold of this method.

Regarding the resolution limit, the most difficult analytes would be structural isomers. However, based on our demonstrated results, structural isomers such as two saccharide isomers (mannose and galactose) and two amino acid isomers (leucine and isoleucine) all reported thoroughly distinguishable event features (Figure RL 1.1), demonstrating a satisfying resolution limit of this technique.
Figure RL 1.1 Clearly distinguish saccharide isomers and amino acid isomers using MspA-FPBA. (a) Top: The chemical structures of D-Mannose (Man) and D-Galactose (Gal). Man and Gal are isomers with identical mass. Bottom: Representative events of Man and Gal. IFPBA represents the open pore current of MspA-FPBA. Events caused by Man and Gal are easily discriminable. (b) The event scatter plot of ΔI versus SD generated from Man and Gal events. Man (n = 103) and Gal (n = 82) events were employed to generate the statistics. Though Man and Gal have identical MW, they are fully discriminated by nanopore. (c) Top: The chemical structures of isoleucine (Ile, I) and leucine (Leu, L). Isoleucine and leucine are isomers with identical mass. Bottom: Representative events of isoleucine and leucine. IFPBA represents the open pore current of MspA-FPBA. Events caused by isoleucine and leucine are easily identifiable. (d) The event scatter plot of ΔI versus SD generated from isoleucine and leucine events. I (n = 237) and L (n = 371) events were employed to generate the statistics. Though

isoleucine and leucine has indistinguishable MW, they are fully discriminated by MspA-FPBA.

Regarding the detection threshold, we believe the reviewer is most likely referring to the limit of detection (LOD). Under idealized measurement conditions, the theoretical sensitivity of single-molecule approaches corresponds to the detection of individual molecules. However, for ultra-low analyte concentration or the presence of only a single moleulce, this may result in an infinite long acquisition time to enable an effective analyte-pore interaction.

Thus, to be realistic, we tentatively define the practical limit of detection (LOD) as the minimum concentration that yields at least 5 binding events during a 10-minute continuous acquisition period. Based on this criterion, our method achieves the LOD values for all 40 analytes. These results were summarized in the Tables RL 1.1-1.4, with LOD values in the micromolar range.
However, it is important to emphasize that these findings were acquired under suboptimal measurement conditions characterized by a large-volume measurement chamber, a single-nanopore sensor configuration, and limited acquisition duration. As an industrial benchmark, Oxford Nanopore Technologies' MinION sequencing platform exemplifies optimized nanoscale detection through its innovative design features. Specifically, an ultra-flat, miniaturized fluidic chamber integrating a high-density array of sensors (~5000). With this industrially validated platform, continuous 24-hour measurement periods could generate significantly enhanced event accumulation, theoretically enabling improvements in sensitivity by several orders of magnitude. The detection limit demonstrated in this study, however, must be considered preliminary due to two critical constraints: (1) the academic proof-of-concept nature of this investigation, and (2) the unavailability of MinION-like devices for experimental use with custom developed pores such as MspA-FPBA.

Table RL 1.1 The limit of detection (LOD) for amino acids. In separate measurements, different amino acids were respectively added to cis. Nanopore measurements were carried out in a 1.5 M KCl buffer (cis: 1.5 M KCl, 10 mM CHES, pH 9.0; trans: 1.5 M KCl, 10 mM MES, pH 6.0). A transmembrane voltage of +160 mV was continually applied during the measurements.

Amino acids	LOD (μM)
A	20
D	10
E	25
F	15
G	10
H	5
I	5
K	15
L	10
M	15
N	10
P	20
Q	10
R	5
S	20
T	10
U	10
V	15
W	10
Y	5
C	10
Ac-K	20
Me-R	20
P-Y	15

Table RL 1.2 The limit of detection (LOD) for NMPs. In separate measurements, different NMPs were respectively added to cis. Nanopore measurements were carried out in a 1.5 M KCl buffer (cis: 1.5 M KCl, 10 mM CHES, pH 9.0; trans: 1.5 M KCl, 10 mM MES, pH 6.0). A transmembrane voltage of +160 mV was continually applied during the measurements.

NMPs	LOD (μM)
AMP	10
UMP	15
CMP	20
GMP	30

ψ 20
m6 A 20
IMP 20

Table RL 1.3 The limit of detection (LOD) for monosaccharides. In separate measurements, different monosaccharides were respectively added to cis. Nanopore measurements were carried out in a 1.5 M KCl buffer (cis: 1.5 M KCl, 10 mM CHES, pH 9.0; trans: 1.5 M KCl, 10 mM MES, pH 6.0). A transmembrane voltage of +160 mV was continually applied during the measurements.

Monosaccharides	LOD (μM)
Ara	250
Fru	150
IdoA	100
GlcNAc	200

Table RL 1.4 The limit of detection (LOD) for peptides. In separate measurements, different peptides were respectively added to cis. Nanopore measurements were carried out in a 1.5 M KCl buffer (cis: 1.5 M KCl, 10 mM CHES, pH 9.0; trans: 1.5 M KCl, 10 mM MES, pH 6.0). A transmembrane voltage of +160 mV was continually applied during the measurements.

Peptides	LOD (μM)
IK	4
LRG	2
SLR	6
GYIK	5
LRGY	1

In real-world biotechnological applications, resolution, versatility, and detection limits (LOD) are all critical, among which resolution is particularly crucial because multiple analytes are simultaneously detected as complex mixtures. The exceptional versatility of MspA-FPBA also enables concurrent detection of diverse analyte types within a single assay configuration, a key advantage for real-world sample analysis, as demonstrated in the yeast extract analysis. Crucially, a high-resolution sensing system ensures accurate and unambiguous identification of each analyte type, preventing misinterpretation errors that could arise from analytical cross-talk or quantification inaccuracies.

To fully address your insightful comments, in the revised manuscript, the LOD at the current measurement setting has been clearly summarized. Some additional discussions regarding optimization of LOD and the advantages of MspA-FPBA in real biological settings have been included. For the ease of your reference, these added texts are also pasted below.

“Although nanopore sensing theoretically offers single-molecule detection limits, this would require infinitely long measurement durations. To establish practical benchmarks, we defined the limit of detection (LOD) as the minimum analyte concentration capable of generating at least five binding events within a 10-minute continuous acquisition period. Based on this criterion, the LODs for amino acids, nucleoside monophosphates (NMPs), saccharides, and peptides are summarized in Supplementary Tables 13-16. Notably, future sensitivity improvements could be achieved through the integration of ultra-flat, miniaturized fluidic chambers with high-density sensor arrays, combined with extended measurement durations similar to a MinION sequencing device.”

“The exceptional versatility of MspA-FPBA enables concurrent detection of diverse analytes within a single assay, offering a key advantage for real-world sample analysis, as demonstrated in the yeast extract analysis. Additionally, its high-resolution sensing ensures accurate and unambiguous identification, preventing misinterpretation errors from analytical cross-talk or quantification inaccuracies.”

Review Comments:

There is no indication of the filter used to produce the electrical traces in each figure. The methods mention that events shorter than 2 ms were removed, but was additional filtering of the data conducted to produce these plots, as they are relatively clean traces? For example, in Figure 6, I expected noisier traces, given it was conducted in yeast extract. If there was additional filtering, then these traces would be more obvious to me. I imagine that other components of the extract could clog the pore. Was there additional data

processing or experimental handling required to maintain the detection resolution? If so, those technical details and discussions need to be added.

Response:

We sincerely acknowledge reviewer 1 for the insightful discussion regarding the current traces. Notably, all the traces presented in the manuscript were not subjected to any further digitial filtering. It should be clarified that the exclusion of events shorter than 2 ms refers to consistently ignoring short residing events (<2 ms) prior to the statistical analysis, rather than removing these events directly from the current trace through digitial filtering. We sincerely apologize for the misunderstanding caused to the reviewer due to the lack of detailed explanations. To clarify this, in the revised manuscript, we have added a detailed explanation regarding the exclusion of events shorter than 2 ms.
The relatively clean current traces observed in the manuscript's figures can be attributed to the following factors. For small molecular analytes, the MspA-FPBA sensor demonstrates selectivity through the incorporation of the FPBA adapter, which specifically responds to target molecules containing primary amine or cis-diol functional groups. Non-complementary small molecules typically yield spiky current events or no detectable signals at all.

Regarding macromolecular analytes, as demonstrated by the yeast cell extract results in Figure 4, ultrafiltration pretreatment effectively removes these species prior to measurement (Figure RL1.2). While nanopore analysis of untreated yeast extracts is technically feasible, macromolecular clogging/trapping events can more or less interrupt the acquisition. Notably, these interference events exhibit distinct signature profiles compared to those of small molecular analytes. The resulting clogging events can subsequently be excluded from statistical analysis during data processing (Figure RL1.3).

In the original manuscript, we mentioned the ultrafiltration step in the methodology session:

“To start with, commercially available yeast cell extract was dissolved in ultrapure water and subjected to ultrafiltration.”

To better describe this, in the revised manuscript in the corresponding figure legend, we have added technical details regarding the ultrafiltration treatment and discussed the necessity of this additional step, supported by the control experiment without ultrafiltration.

To fully address your comment, results of events prior to and after the ultrafiltration treatment have also been added to the revised manuscript along with some relevant discussions. For the ease of your reference, these added texts are also pasted below.

“Notably, all traces presented in this study were not subjected to any further digital filtering. Short-residing events (<2 ms) were consistently excluded prior to the statistical analysis.”

“Ultrafiltration pretreatment effectively removes macromolecular analytes from yeast cell extract (Supplementary Fig. 48). Specifically, the extract solution was loaded into an ultrafiltration tube with a 3 kDa molecular weight cut-off and ultrafiltered at 6,000 r.p.m. for 60 min at 4 °C. Although nanopore analysis of untreated yeast cell extract is also feasible, macromolecular clogging/trapping events can more or less interrupt data acquisition (Supplementary Fig. 49). However, these interference events exhibit signature profiles distinct from those of small molecular analytes and can be excluded from statistical analysis after data acquisition.”
Figure RL 1.2 The extended representative current trace of yeast cell extract after ultrafiltration. The yeast cell extract was dissolved in ultrapure water with a concentration of 100 mg mL-1. Afterwards, the yeast cell extract solution was loaded into an ultracentrifuge tube with a 3 kDa molecular weight cut-off. The filtration was then performed at 6,000 r.p.m. for 60 min at 4 °C. Then, the pH of the ultrafiltrate was adjusted to 7.0 by titration with KOH. To initiate the nanopore measurement, 5 μL of the above filtrate was added to the cis side and a +160 mV bias was continuously applied. Finally, various nanopore events were classified and identified using machine learning. For clarity, IFPBA was marked by a gray dashed line. The transient spike signals similar to nanopore blockage were marked with gray arrows. Events were automatically classified via the few-shot learning (Fig. 3) and labeled with color-coded dots (amino acids: orange; NMPs: blue; saccharides: green; peptides: purple). With the trained bagged trees model (Fig. 2a), different amino acids and NMPs were further identified.

Figure RL 1.3 Analysis of yeast cell extract without ultrafiltration. The yeast cell extract was dissolved in

ultrapure water to a concentration of 100 mg mL-1, without being subjected to ultrafiltration. Then, the pH of the solution was adjusted to 7.0 by titration with KOH. To initiate the nanopore measurement, 5 μL of the above solution without ultrafiltration was added to the cis side and a +160 mV bias was continuously applied. Subsequently, frequent nanopore blockages were observed. The measurement was eventually terminated due to severe nanopore clogging. Nanopore clogging can be effectively removed by flipping of the applied voltage. For clarity, IFPBA was marked by a gray dashed line, and the nanopore blockages were labeled with gray dots.
Review Comments:

It would be helpful to include an explanation of how the parameters SD and ΔI were chosen as the key identifying elements. The authors list all the parameters analyzed and provide them in supplementary tables, but a more detailed explanation of whether other parameters could also be used to distinguish these analytes would be beneficial.

Response:

We sincerely appreciate Reviewer 1's valuable suggestions regarding parameter selection. It should however be clarified that ΔI and SD were never utilized as only parameters for event predication. Instead, they served only as parameters for 2D data visualization. For all analytical predictions, by utilization of a custom machine learning model, we simultaneously employed a comprehensive set of nine parameters (ΔI, SD, LocalSD Ratio, min, max, range, toff, skew, and kurtosis) for event identification.

The reviewer is correct in noting that ΔI and SD form an effective visualization pair. As illustrated in Figure RL 1.4, by using amino acids S, A, Q, T, V, M, I, L, and W as model systems, the ΔI vs. SD scatter plot (marked with a red box) successfully resolves nine distinct event clusters corresponding to different analytes. This contrast sharply with scatter plots employing alternative parameter combinations (shown in other panels), which exhibit significant cluster overlaps, demonstrating the superior 2D visualization capability of the ΔI/SD parameter pair.

The inquiry from Reviewer 1 also motivated us to systematically assess the contribution of distinct features to the model's performance. To accomplish this, we employed the Permutation Feature Importance method to calculate ImportanceMean values for various event parameters. Permutation Feature Importance quantitatively evaluates each feature's contribution to the model's analytical discrimination capability. Specifically, higher ImportanceMean values directly correlate with greater feature importance, indicating a more pronounced influence on the model's ability to differentiate analytes. As shown in Table RL 1.5 and Figure RL 1.5, the ImportanceMean values of ΔI and SD reached 0.8684 and 0.527, respectively, indicating that they made the greatest overall contribution to the model's ability to distinguish all analytes. The ImportanceMean values of the other parameters, including LocalSD Ratio, toff, max, min, range, kurt, and skew, were 0.1163, 0.0617, 0.0115, 0.0064, 0.004, 0.0029, and 0.0013, respectively, showing that their overall contribution to the model's ability to distinguish all analytes decreased in this order. Overall, all nine event parameters simultaneously contribute to the model's identification capacity. However, ΔI and SD demonstrate a more pronounced contribution, as also noted by the reviewer.

Figure RL 1.4 2D scatter plots of different event parameters. Single-molecule events of the nine amino acids (S, A, Q, T, V, M, I, L, and W) were analyzed using event parameters including ΔI, SD, LocalSD Ratio, min, max, range, toff, skew, and kurt. In the event scatter plot of ΔI versus SD (marked by a red box), the 2D visualization distinctly separates nine event clusters, with each cluster corresponding to a specific amino acid type. In contrast, scatter plots using other parameter combinations demonstrate significant cluster overlap between different amino acid types. These results indicate that the ΔI and SD parameter pair provides optimal 2D discrimination capability for single-molecule sensing analysis. Each visualization was generated from 1396 events to ensure statistical significance.

Table RL 1.5 The ImportanceMean values of different event parameters on the model's accuracy. The ImportanceMean values of nine event parameters (ΔI, SD, LocalSD Ratio, min, max, range, toff, skew and kurt) were calculated for the Bagged Trees model used to distinguish all 40 analytes, using the Permutation Feature Importance method. ImportanceStandardDeviation represents the standard deviation of the importance of the event parameters.

Parameters	ImportanceMean	ImportanceStandardDeviation
ΔI	0.8684	0.0045
SD	0.5270	0.0044
LocalSD Ratio	0.1163	0.0020
toff	0.0617	0.0028
max	0.0115	0.0012
min	0.0064	0.0010
range	0.0040	8.2033×10-4

kurt 0.0029 5.8286×10-4
skew 0.0013 2.5816×10-4

In the revised manuscript, we have comprehensively addressed this concern by incorporating additional discussions on the topic. We sincerely appreciate Reviewer 1 for raising this insightful question, which directly inspired our implementation of feature importance analysis. The inclusion of this investigation has significantly enhanced our understanding, allowing for a more precise characterization of the distinct contributions of various event features compared to the previous version of the manuscript. For the ease of your reference, these added texts are also pasted below.

Figure RL 1.5 Visualization of the ImportanceMean values for different event parameters. A larger ImportanceMean value indicates that the parameter is more important in the model's differentiation of all analytes and has a greater overall contribution. The ImportanceMean values of nine event parameters (ΔI, SD, LocalSD Ratio, min, max, range, toff, skew and kurt) were calculated for the Bagged Trees model used to distinguish all 40 analytes, using the Permutation Feature Importance method.

“To visualize nanopore sensing results, scatter plots based on different parameter combinations were constructed using amino acids S, A, Q, T, V, M, I, L, and W as model systems (Supplementary Fig. 17). Among them, the ΔI vs. SD plot clearly resolved nine distinct event clusters corresponding to different analytes. Therefore, ΔI and SD were used for 2D data visualization throughout this study.”

“For all analytical predictions, a comprehensive set of nine parameters (ΔI, SD, LocalSD Ratio, min, max, range, toff, skew, and kurt) was simultaneously employed for event identification. To systematically assess the contribution of distinct features, the Permutation Feature Importance characterization was employed49, 50. Specifically, higher ImportanceMean values directly correlate with greater feature importance. As shown in Supplementary Table 8 and Supplementary Fig. 29, the ΔI and SD has made the greatest contribution to the model's discrimination capacity, followed by LocalSD Ratio and toff. Other parameters, including max, min, range, kurt, and skew also show their overall contribution. Finally, all nine event parameters simultaneously contribute to the model's identification capacity.”

Review Comments:

The overall approach should definitely be benchmarked against other competing technologies. For example, in the glycosylated peptide experiments, additional enzymatic degradation of a non-glycosylated peptide would be necessary. Alternatively, the authors could explain or provide additional data to demonstrate their confidence that the three unknown events are not artifacts of enzymatic fragmentation.

Response:

We acknowledge reviewer 1 for the constructive suggestion to benchmark our approach against other competing techniques. Strictly following the reviewer’s suggestion, we have benchmarked our approach with mass spectrometry. Additionally, we performed enzymatic digestion of non-glycosylated peptides followed by nanopore measurements.

For the glycosylated peptide GMQRN(GlcNAc)IS, as shown in Figure RL 1.6, the mass spectrometric results revealed molecular ion peaks corresponding to G, M, Q, R, and N(GlcNAc)IS. The mass spectrometry results were consistent with the nanopore measurement results obtained using our approach (Extended Data Fig. 7).

In addition, we performed nanopore measurements on the enzymatic digestion filtrate of the non-glycosylated peptide. With the previously trained bagged trees model, it was confirmed that all seven clusters were generated by G, M, Q, R, N, I, and S (Figure RL 1.7a-b). The three unidentified events (abbreviated as UE1, UE2, UE3) observed in the glycosylated peptide experiment (Figure RL 1.7c-d) were not detected anymore. This indicates that the three unidentified events are not artifacts of the enzymatic degradation. Combined with the above mass spectrometry results (Figure RL 1.6), the three identified events were further confirmed to be generated by N(GlcNAc)IS.

Figure RL 1.6 Mass spectrometric characterization of the enzymatic digest filtrate of the glycosylated peptide GMQRN(GlcNAc)IS. Molecular ion peaks corresponding to G (m/z = 76.0393), M (m/z = 150.0597), Q (m/z = 169.0597), R (m/z = 158.0919), and N(GlcNAc)IS (m/z = 536.2553) were observed, consistent with the nanopore measurement results in Extended Data Fig. 7.

In the revised manuscript, we have added mass spectrometric characterization and nanopore measurements of the non-glycosylated peptide digestion filtrate. Together with the mass spectrometry results and the non-glycosylated peptide control experiment, it confirmed that the three unidentified events were produced by the N(GlcNAc)IS fragment generated during the glycopeptide enzymatic digestion. For the ease of your reference, these added texts are also pasted below.

“The corresponding mass spectrometric results confirmed the generation of G, M, Q, R, and N(GlcNAc)IS (Supplementary Fig. 42). Nanopore measurements on the enzymatic digestion filtrate of the non-glycosylated peptide were also performed and the detected events were identified as G, M, Q, R, N, I, and S (Supplementary Fig. 43 and Supplementary Fig. 44). The three previously unidentified events were not detected. Together with the mass spectrometry results and the results acquired with the non-glycosylated peptide, it is confirmed that the three unidentified events indeed originated from the N(GlcNAc)IS fragment generated during the glycopeptide enzymatic digestion.”
Figure RL 1.7 Nanopore measurements of the enzymatic digestion filtrates of non-glycosylated and glycosylated peptides. a, Representative traces of the heptapeptide with the sequence GMQRNIS following LAP enzymatic digestion (Methods). Representative events corresponding to different amino acids were automatically predicted via machine learning and labeled with color-coded dots (G: dark red; M: dark purple; Q: light green; R: orange; N: brown; I: deep rose; S: muted mauve). b, The scatter plot of ΔI versus SD for the amino acid events obtained from a. The scatter plot was generated using a 20 min continually recorded trace. The seven clusters of amino acids corresponding to G, M, Q, R, N, I and S were identified by machine learning. The three unidentified events (UE1, UE2, and UE3) were not detected, indicating that they were not artifacts. (c) The representative traces and (d) the scatter plot of the glycosylated peptide with the sequence GMQRN(GlcNAc)IS following LAP digestion and ultrafiltration. The amino acid events were automatically predicted via machine learning and labeled with color-coded dots (G: dark red; M: dark purple; Q: light green; R: orange). UE1, UE2 and UE3 are indicated with red dashed boxes. Combined with the above mass spectrometry results (Figure RL 1.6), the three unidentified events were indeed generated by N(GlcNAc)IS.

Review Comments:

The first sentence of the abstract is the same as the first sentence of the main text.

Response:

We thank reviewer 1 for this wording suggestion. Strictly following your request, in the revised manuscript, we have revised the opening sentences to eliminate redundancy. For the ease of your reference, the revised opening sentences of the abstract and the main text are provided below, respectively.

“Cellular functions arise from the coordinated interactions of key molecules such as proteins, RNA, and glycans.”

“Proteins, RNAs, and glycans constitute three fundamental molecular classes that collectively regulate cellular function.”
Reviewer #2:

Review Comments:

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Biotechnology initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Response:

We sincerely appreciate reviewer 2's insightful suggestions and valuable time. The questions raised by you and your co-reviewer are greatly appreciated and have all been well addressed in a point-by-point manner. At anytime, please never hesitate to let us know if there is any remaining unclear issues, we will be more than happy to clarify it.
Reviewer #3:

Review Comments:

The manuscript by Yao et al focuses on analysis of various small-molecule agents using a clever pore modification. The authors have modified an MspA nanopore with a formyl-phenylboronic acid moiety to form a pore they call "MspA-FPBA". This engineered pore, which contains both a boronic acid moiety and an aldehyde residue, was then used to detect all of the 21 canonical amino acids, as well as the RNA mononucleoside monophosphates (NMPs), and some sugars. The basis of these detections are that: 1) α-amino acids interact with aldehydes by forming a Schiff Base with the α-amino groups, which results in a long dwell time of interaction (~1 s). 2) NMPs interact via reversible boronic acid/diol interactions, and 3) sugars interact the same way (through diol/boronic acid interactions). Altogether, almost 40 molecules are detected and distinguished using this scheme (among these molecules are short peptides). For such small molecules (MW < 1 kD), 1 second long interaction times with the pores is extremely long - in fact, one would expect to find unobservable sub-μs long dwell times. The reversible chemical interaction, as well as the careful positioning of the anchoring groups in the pore constriction, are responsible for the observations of unique ion current signatures.

Response:

We extend our sincere gratitude to Reviewer 3 for accurately describing the innovative capabilities of this newly engineered MspA-FPBA nanopore platform. The hetero-octameric MspA pore features a strategically positioned formyl-phenylboronic acid (FPBA) adapter at the constriction region, enabling the unified sensing capacity shown in this manuscript. This functional adapter, which has never been previously reported, was custom synthesized by us. It incorporates dual chemical recognition moieties, a boronic acid for diol detection and an aldehyde group for amino functionalities, enabling simultaneous analysis of 40 distinct small molecules. We particularly appreciate Reviewer 3's comprehensive understanding of our detection mechanisms.

For amino acids and peptides, the α-amino groups specifically engage with the aldehyde functionality through iminoboration reactions, forming boronic acid-stabilized Schiff base complexes. This interaction results in characteristic dwell times exceeding one second. Meanwhile, the cis-diol moieties present in nucleoside monophosphates and carbohydrates undergo boronic ester formation with the adapter's boronic acid component. As noted by Reviewer 3, these molecular interactions (for compounds with molecular weight < 1 kDa) generate prolonged current signatures rather than transient sub-millisecond events, allowing for reporting of rich sensing information and subsequent high resolution event identifications. This sensing capacity is also verified with complex mixtures of biological samples to show its validity.
Review Comments:

Full disclaimer: this paper is a very impressive demonstration of the usability of nanopores for discrimination among various small molecule species. The writing is impeccable, and the results are clearly presented. I have no comments on the rigor and writing.

Response:

We greatly appreciate reviewer 3 for considering this work very impressive in distinguishing a large varity of small-molecule species. We also thank reviewer 3 for acknowledging the rigor of our study and the quality of the writing.

Review Comments:

Having said that, however, I must admit that I have a hard time with its positioning in the realm of usability and novelty for Nature Biotechnology. One major issue here is that all of these molecules are detected at μM to mM concentrations, and mass spectrometers can equally do a good job at detecting and quantifying these molecules and more in a mixture.

Response:

We highly appreciate the comments raised by reviewer 3 regarding a technical comparison with MS and the sensitivity of this technique. They have both been separately addressed below.

1. Comparison with MS

We agree with reviewer 3 that mass spectrometry is a well-established technique for the detection and quantification of diverse molecules in complex mixtures. However, conventional MS still faces inherent challenges in distinguishing structural isomers with identical masses (Nature Reviews Methods Primers 2, 48 (2022). https://doi.org/10.1038/s43586-022-00128-4; Analytical Chemistry 88, 10757-10766 (2016). https://doi.org/10.1021/acs.analchem.6b03409), such as amino acid isomers leucine and isoleucine, as well as saccharide isomers galactose and mannose (Figure RL 3.1). These isomers are however directly discriminable by this nanopore sensor when tested in a side by side comparison (Figure RL 3.2). Although tandem MS can distinguish constitutional isomers via characteristic product ions, it generally cannot directly differentiate chiral enantiomers (e.g., D- and L-amino acids) due to lack of diagnostic fragments (Analytical Chemistry 95, 6996−7005 (2023). https://doi.org/10.1021/acs.analchem.3c00495). Therefore, chiral separation, derivatization, or chiral reagents are usually needed to convert them into diastereomers for MS/MS differentiation. However, in this work, the developed MspA-FPBA can clearly distinguish saccharide isomers and amino acid isomers (Figure RL 3.2), without any sophsicated operation.

Another significant advantage of nanopore lies in its exceptionally rapid, economical, and portable sensing capabilities. This combination of features enables point-of-use implementation by untrained personnel in diverse settings without requiring extensive specialized training or investment in sophisticated, costly laboratory equipment. However, the previously demonstrated nanopores are rather specialized for a single type of analyte, demaning the development of a more unified nanopore sensor for the detection of a wide range of analtyes all in the same assay. However, the implementation of a single, highly integrated sensor would not only reduce hardware component cost and size but also significantly simplify the overall system operation complexity, extremely suitable for application scenarios that MS is not accessible or required.

Figure RL 3.1 Saccharide and amino acid isomers resolved with mass spectrum. (a) The chemical structures of D-Mannose (Man) and D-Galactose (Gal). Man and Gal are isomers. (b-c) MS spectra of (b) Man and (c) Gal. Their molecular ion peaks are exactly the same. (d) The chemical structures of isoleucine (Ile, I) and leucine (Leu, L). Ile and Leu are isomers with identical mass. (e-f) MS spectra of (e) Ile and (f) Leu. Their molecular ion peaks are exactly the same.

Figure RL 3.2 Clearly distinguish saccharide isomers and amino acid isomers using MspA-FPBA. (a) Top: The chemical structures of D-Mannose (Man) and D-Galactose (Gal). Man and Gal are isomers with identical mass. Bottom: Representative events of Man and Gal. IFPBA represents the open pore current of MspA-FPBA. Events caused by Man and Gal are easily discriminable. (b) The event scatter plot of ΔI versus SD generated from Man and Gal events. Man (n = 103) and Gal (n = 82) events were employed to generate the statistics. Though Man and Gal have identical MW, they are fully discriminated by nanopore. (c) Top: The chemical structures of isoleucine (Ile, I) and leucine (Leu, L). Isoleucine and leucine are isomers with identical mass. Bottom: Representative events of isoleucine and leucine. IFPBA represents the open pore current of MspA-FPBA. Events caused by isoleucine and leucine are easily identifiable. (d) The event scatter plot of ΔI versus SD generated from isoleucine and leucine events. I (n = 237) and L (n = 371) events were employed to generate the statistics. Though isoleucine and leucine has indistinguishable MW, they are fully discriminated by nanopore.

2. The sensitivity and potential engieering optimization

In response to the reviewer's comment on the sensitvity, it is important to note that for a nanopore system, the theoretical limit of detection (LOD) corresponds to a single molecule. However, in practice, achieving reliable detection of a true single molecule may require an impractically long duration for successful analyte-pore interaction. To address this, in this study, we have tentatively defined the LOD as the minimum analyte concentration yielding at least 5 binding events within 10 minutes of continuous data acquisition. In the revised manuscript, we systematically evaluated the LOD, and the results consistently fall in the micromolar (μM) range, consistent with the reviewer's observation.

However, it is crucial to emphasize that these findings were obtained under suboptimal experimental conditions and are intended solely as a proof-of-concept, academic demonstration. Further industrial optimization, such as employing a small-volume chamber, integrating a high-density nanopore array, and extending acquisition times, could significantly enhance the sensitivity performance. For context, the MinION sequencing platform, an industrially validated system, demonstrates a desired hardware suitable for

prolonged measurements using an array to accumulate events to reach a sensitivity a few orders of magnitude better than the current demonstration. However, as an academic group, we lack access to such industrial-grade setups simultanouesly without any prior installed pores for an sensitivity demonstration. However, a future collaboration with pioneering companies such as ONT may achieve this demonstration.

We are gratified that the reviewer acknowledges the impressive performance of the MspA-FPBA system. While substantial technical refinement could surely enhance the system's sensitivity by several orders of magnitude, innovation in this domain appears to be of lower priority at present.

Review Comments:

Second, the authors have already demonstrated the detection of amino acids as well as sugars and NTPs in previous papers. So, this paper feels a bit like an incremental achievement to the previous works by the authors (all of which I genuinely admire).

Response:

We sincerely appreciate Reviewer 3's recognition of the high quality of our prior work. While we acknowledge the reviewer's perspective, we respectfully differ regarding the assessment of MspA-FPBA development as merely incremental. It simply reminded me of the iPhone's groundbreaking introduction. When Steve Jobs unveiled the iPhone, he ingeniously integrated telephone, internet browser, camera, MP3/MP4 player, and calculator functionalities within a single, cohesive architecture. Although these functions existed as separate devices beforehand, the iPhone created a revolutionary new era by enabling unprecedented application ecosystems. This parallels our current work, which similarly follows an innovative integration strategy.

Below, we present key technical features underscoring its novel contributions and indispensible role in nanopore sensing:

First Synthesis of Maleimido-C2-FPBA and Its Validation. Our work represents the inaugural report of the maleimido-C2-FPBA compound, a core innovation that enables unified nanopore sensing. This compound was specifically designed and synthesized through extensive screening efforts, with comprehensive synthetic protocols and characterization data provided (Methods and Supplementary Figs. 1-4). The resulting nanopore signals corroborate its validity as a foundational component of our platform.

Platform for Further Expansive Analytical Scope. The MspA-FPBA system inherently supports detection of other diverse analytes: including but not limited to D-amino acids, D-peptides, nucleotide sugars, oligosaccharides, and oligonucleotides, without assay-specific alterations. This unified sensing compatibility, unmatched by existing nanopore sensors, establishes a paradigm shift in single-assay multiplexed analysis.

Native Biological Sample Analysis without Pre-Targeting. In the nanopore field, current systems require prior knowledge of target molecules to analyze complex biological matrices. This platform, however, enables holistic profiling of coexisting biomolecules (e.g., amino acids, NMPs, saccharides, peptides) in natural samples, as demonstrated through yeast extract analysis. This represents a significant advancement in comprehensive biomolecular profiling.

Characterization of Hybrid Biomolecular Complexes. The system's architecture is uniquely suited for analyzing structurally complex species such as glycopeptides, glycoRNAs, and RNA-binding peptides, which integrate features of glycans, nucleic acids, and peptides in the same molecule. New data in the revised manuscript (e.g., G2 peptide analysis) exemplify this capability, addressing hybrid molecule characterization with unprecedented specificity.

Enabling Next-Generation Multi-Omics Sequencing. The platform lays groundwork for future enzymatic sequencing architectures. Possibly by integrating proteases, nucleases, and glycosidases at the nanopore entrance, sequential degradation of proteins, nucleic acids, and glycans might yield sequence/structural information through time elapsed recording analysis. Such multi-omics approaches for glycoproteins, glycoRNAs, and RNA-protein complexes would require a unified sensing capacity that MspA-FPBA is demonstrating. When sequencing glycoRNAs or glycoproteins, nanopores must be configured to be responsive of nucleotides, saccharides, or amino acids upon their stochastic appearance. Notably, such a

sequencing application inherently precludes the possibility of dynamic pore-type switching during sequencing.

Action without the reliance of metal ions. Conventional nanopore recognition of amino acids typically requires pore modification via metal-ion coordination. However, this approach often introduces conflicting requirements when multiple enzymatic components are also involved in the system. For instance, DNA polymerase activity generally necessitates Mg^{2+} ions, while certain proteolytic enzymes require Zn^{2+}, Co^{2+} or Cu^{2+} for optimal function. The MspA-FPBA system eliminates these metal-ion dependencies, thereby circumventing the need for metal-ion-mediated pore modification. This design simplifies the experimental setup and mitigates potential incompatibilities between different metal-ion requirements within the system.

These contributions collectively establish MspA-FPBA not as an incremental advance, but as a transformative platform with broad implications for biomolecular analysis and future sequencing technologies.

Review Comments:

Therefore, while I have no technical comments about the data or conclusions made in the manuscript, I believe that this work is more suitable when adjoined to a more tangible application that would highlight the importance of detecting all of these species (or some of them) in a complex mixture. The yeast extract data in figure 6 is interesting, but falls short of quantification and exhaustive analysis that would highlight anything particularly useful to the bioanalytical field.

Response:

We sincerely appreciate reviewer 3's insightful comments and suggestions regarding the need to associate this approach with a more tangible application. The original aim of developing this unified nanopore sensor stems from the inherent complexity of real-world biological analysis, where the components of a system are often unknown. For instance, the system may consist of mixed combinations of amino acids, peptides, and saccharides, such as demonstrated in the yeast extract example provided in the original manuscript. Furthermore, it may involve hybrid biomacromolecules formed by these components, including glycoRNA or glycopeptides, which also requires a unified nanopore sensor for efficient component analysis. To highlight the importance of the unique capability of MspA-FPBA to enable unified detection of multiple analytes and generate distinct event signatures for different analyte types, in the revised manuscript, we furether demonstrated its applicability in glycopeptide analysis.

Owing to the presence of isomeric monosaccharides and amino acids, current MS-based glycoproteomic approaches still face challenges in the direct compositional identification and quantification of native glycopeptides (Figure RL 3.1). To demonstrate the capability of MspA-FPBA for native glycopeptide compositional analysis, the G2 glycopeptide, a naturally occurring complex-type N-glycopeptide derived from egg yolk, was selected as an ideal model sample. To start with, the commercially available G2 glycopeptide was subjected to hydrolysis to obtain its compositional monomers. Subsequently, the hydrolysis products were added to the cis chamber for nanopore measurements. With the aid of few-shot learning, compositional identification and quantification of the G2 glycopeptide were successfully achieved (Figure RL 3.3).

Besides glycopeptides, the MspA-FPBA developed in this work is also theoretically applicable to the compositional analysis and quantification of other critical hybrid biomolecules, such as glycoRNAs and RNA-binding peptides (Figure RL 3.4). However, to date, no competing nanopore-based approaches reported by other groups have demonstrated similar versatility or sensitivity.

Figure RL 3.3 Identification and quantification of amino acid and saccharide components in G2 glycopeptide. Experimentally, 3 mg of G2 glycopeptide was dissolved in 250 μL 0.5 M TFA and incubated at 80°C for 12 h. Then, sodium phosphate was added to the acid hydrolysis product to a final concentration of 50 mM, and the pH was adjusted to 8.0. Subsequently, the solution was digested with LAP and ultrafiltered. With MspA-FPBA, 10 μL of the above filtrate was added to the cis side and a +160 mV bias was continually applied. Finally, the components of the G2 glycopeptide were identified and quantified by nanopore analysis.

To fully address your inquires, some additional results and relevant discussions have been added in the revised manuscript. They are also pasted below, for the ease of your reference.

“Due to the presence of structural isomers, current mass spectrometry-based approaches encounter significant challenges in achieving direct compositional identification and quantification of native glycopeptides (Supplementary Fig. 52). In contrast, the MspA-FPBA platform demonstrates distinct advantages in isomer discrimination (Supplementary Figs. 19, 53). To demonstrate the technical potential for native glycopeptide analysis, a naturally occurring complex-type N-glycopeptide (G2) isolated from egg yolk was selected as a model system and subjected to hydrolysis for nanopore characterization (Methods and Fig. 5a). The raw traces are presented in Fig. 5b-c, accompanied by corresponding scatter plot analysis in Fig. 5d and Supplementary Fig. 54. The unidentified events were classified as saccharides by few-shot learning, suggesting that they might be Gal and Man, two core monosaccharide components in mammalian N-linked glycans68. This assignment was further validated through direct comparison with reference mass spectrometric profiles of standard Gal and Man (Supplementary Fig. 53). As shown in the quantification results (Fig. 5e and Supplementary Table 18), despite minor experimental discrepancies, the nanopore-derived measurements maintain a high degree of consistency with reference values. These findings confirm the suitability of this nanopore platform for both compositional profiling and quantitative analysis of native glycopeptides.

Beyond glycopeptides, MspA-FPBA demonstrates theoretical applicability for compositional profiling and quantitative analysis of other hybrid biomolecules, including but not limited to glycoRNAs and RNA-binding peptides, utilizing the versatility of the maleimido-C2-FPBA adapter. To the best of our knowledge, no previously reported nanopore-based methods have achieved comparable versatility and resolution which is suitable for complex biomolecular characterization. While its proof-of-concept demonstration with a representative glycopeptide validates the technical feasibility of this methodology, the platform represents a promising breakthrough for integrated single-molecule analysis of coexisting biomolecular entities, an analytical challenge that remains unattainable through individual nanopore technology in the field. Looking forward, this approach could revolutionize structural characterization of distinct molecular domains and offer novel mechanistic insights into cross-omics interactions at the glycome-transcriptome-proteome interface. Moreover, the platform's sensing mechanism suggests its potential to discover previously undetected post-translational modification (PTM) types and epigenetic modifications through synergistic advancements in corresponding databases and machine learning algorithms.”

Figure RL 3.4 Compositional analysis of glycopeptides, glycoRNAs, and RNA-binding peptides. The MspA-FPBA nanopore developed in this work can serve as a unified nanopore sensor for compositional analysis of these native hybrid biomolecules. Bare pore and previously reported engineered nanopore cannot achieve this goal.

Finally, we sincerely thank Reviewer 3 for the insightful suggestion that enhanced our work's quality and scope. As the first study using the unified MspA-FPBA nanopore platform, our findings establish a critical foundation for field advancements. Through collaborative development, this technology might further expand its analytical scope to become the first comprehensive analytical platform for nucleic acids, glycans, and proteins all within a single setup. Currently, nanopore proteomics and glycomics represent both the most prominent and challenging frontiers in the nanopore field. The development of MspA-FPBA technology holds significant promise for advancing these individual research directions, their synergistic integration, and ultimately paves the way for nanopore sequencing of protein and glycan.

We also understand that the power from a junior group like us is limited, we thus meticulously documented the MspA-FPBA sensor’s preparation, envisioning its disclousure may soon attract its wide application within the field and might unlock applications even beyond our current imagination. However, timely dissemination would become important. We are deeply honored by the peer review opportunity offered by Nature Biotechnology. We are now at the threshold of bringing this new capability to the wider scientific community and future researchers who stand to benefit most. At this stage, what stands between this technology and the community it could serve is the support of all reviewers. We sincerely hope that, for the sake of further progress to the field, Reviewer 3 may find our revised manuscript and responses compelling, and may support the publication of this work.Your expert guidance and support to the publication of this work is extremely appreciated.

Reviewer #4:

Review Comments:

The manuscript presents an integration of machine learning with a cleverly modified single-molecule nanopore sensing platform for molecules. It presents a strategy to detect single biomolecule types (amino acids, post-translationally modified amino acid,

nucleotides, epigenetically modified nucleotides, and glycans) as well as the same molecules in mixtures. The apparent strategy would be, in the future, to have a cleaving molecule such as a protease/nuclease/glycoside hydrolase at the entrance of the pore, and use that to sequentially degrade respective molecules to be able to sequence each of these molecules. The current work takes a step in this direction. My co-reviewers and I have the following comments on the manuscript. Our main concern is with the validation of nanopore detected analytes in mixtures from both the nanopore experiments and the machine learning perspectives.

Response:

We sincerely thank Reviewer 4 for recognizing the innovation and the advantage of this newly developed nanopore system. It also represents the first nanopore that demonstrates a high versatility for single-biomolecule detection of amino acids, post-translationally modified amino acids, nucleoside monophosphates (NMPs), epigenetically modified NMPs, carbohydrates, and peptides all in the same assay.

We greatly value Reviewer 4's insight into the potential translational applications of this technology in molecular sequencing. As highlighted by the reviewer, the proposed strategy of immobilizing proteolytic, ribonucleolytic, and glycosidic enzymes at the nanopore entrance to enable sequential degradative analysis of peptides, RNA molecules, and glycans represents a significant and compelling future direction. We are particularly pleased that the reviewer acknowledges our current work as establishing a foundational step toward this important technological goal.

We are grateful for Reviewer 4 and the co-reviewers' constructive critiques, all of which have been addressed with comprehensive experimental validation following your valuable suggestions. Concerning the primary inquires regarding analyte identification in mixed-sample sensing, we have provided rigorous independent verification data that fully satisfy the raised concerns. For clarity, we have organized our detailed responses to all questions in a systematic point-by-point format below.

Review Comments:

Overall, the manuscript is very well done and provides appropriate details. The use of the unified sensing adaptor maleimido-C2-FPBA adaptor on MspA is definitely a significant advance. The demonstrations of single species detection is convincing.
Response:

We are profoundly grateful to Reviewer 4 for your insightful recognition of both the manuscript's comprehensive quality and the methodological rigor in presenting technical details. Traditional nanopore detection paradigms have been inherently limited to identifying singular analyte classes. Our breakthrough lies in the development of a strategically engineered maleimido-C2-FPBA adaptor integrated at the pore constriction, enabling unified identification across multiple analyte categories within a single, multifunctional nanopore system. The development of this universal sensing architecture required extensive experimental optimization and molecular engineering, which we are delighted to see has been appreciated by Reviewer 4, which is sincerely acknowledged.

Review Comments:

While experiments with mixtures provide clear signals in almost all cases that are classifiable, I worry about the “ground truth” in such cases. How do we know which species is going through for sure. We have the nanopore signal which indicates a specific species but there is no independent validation of this in the experiments. A possible way would be to have specific mixtures (varying concentrations of species) and then through significant number of reads through the nanopores try to reproduce the concentration of species. This may or may not be feasible but some demonstration like this is needed. If the above is feasible, this may also be a way of determining the overall sequence of protease treated protein mixtures (for example) by combining experiments with ML training on single proteins solutions and multiple mixtures of proteins of increasing complexity.

Response:

We greatly appreciate Reviewer 4 for the valuable suggestion to employ customized mixtures with varying concentrations of target species to demonstrate nanopore

quantification experiments. Such demonstration would provide independent validation of our machine learning model's species identification in mixed-solution sensing, thereby strengthening the credibility of this method.

In the nanopore field, quantitative analysis is conventionally performed by monitoring the event occurrence rate, which is linearly correlated with the analyte concentration. Specifically, the detection frequency of nanopore events (measurable parameter F, defined as events per minute for a specific analyte when the nanopore is unoccupied) exhibits a concentration-dependent relationship. By taking methionine as a model analyte, maintaining consistent experimental conditions while varying methionine (Met) concentration clearly demonstrates this dependence (Figure RL 4.1).

In the revised manuscript, we selected representative analytes of four analyte classes: amino acids (Met, Asn, Tyr, Arg), nucleoside monophosphates (UMP), saccharides (Fru), and peptides (tripeptide LRG). Their concentration-dependent nanopore calibration profiles are summarized in Supplementary Tables 11-12 and Supplementary Figures 30-36. These data validate the quantitative relationship between analyte concentration and event frequency, supporting the proposed mixed-solution analysis capacity.

Figure RL 4.1 The concentration dependence of Met sensing using MspA-FPBA. The measurements were carried out in a 1.5 M KCl buffer (cis: 1.5 M KCl, 10 mM CHES, pH 9.0; trans: 1.5 M KCl, 10 mM MES, pH 6.0). A transmembrane voltage of +160 mV was continually applied during the measurements. Met was added to cis with a final concentration of 30-600 μM. (a-c) Representative traces acquired with varying Met concentrations. (d) The plot of F versus the Met concentration. The plot is overlaid with the corresponding linear fitting result. Three independent measurements were performed to show data consistency (N = 3). All statistical results are also summarized in Supplementary Table 11. The data in (d) demonstrates mean ± standard deviations derived from results of three independent measurements (N = 3). The error bars represent standard deviation values.

In accordance with the reviewer's suggestion, we prepared a standard mixture sample by selecting Met, Asn, Tyr, Arg, UMP, Fru, and the tripeptide LRG. Each species had distinct, well-defined concentrations. This mixture enabled us to demonstrate the simultaneous quantification of amino acids, nucleoside monophosphates (NMPs), saccharides, and peptides. Specifically, we dissolved approximately 1.7 mg of LRG, 3.5 mg of arginine (Arg, R), 4.0 mg of asparagine (Asn, N), 6.0 mg of methionine (Met, M), 11.3 mg of L-tyrosine disodium salt hydrate (Tyr, Y), 55.2 mg of uridine 5′-monophosphate disodium salt (UMP), and 90.1 mg of D-fructose (Fru) in 10 mL of ultrapure water. This yielded final concentrations of 0.5, 2.0, 3.0, 4.0, 5.0, 15.0, and 50.0 mM, respectively. A 10 μL aliquot of this mixture was analyzed using MspA-FPBA. As shown in Figure RL 4.2, all seven species were effectively detected by the nanopore. The quantitative results for each component showed strong agreement with their known concentrations (Table RL 4.1), confirming the capability of our method to simultaneously quantify amino acids, NMPs, saccharides, and peptides in complex mixtures.

Figure RL 4.2 Quantitative analysis of a mixed sample containing Met (M), Asn (N), Tyr (Y), Arg (R), UMP, Fru, and the tripeptide LRG. (a-b) Continuous traces acquired with a mxied sample containing M, N, Y, R, UMP, Fru, and LRG. 1.7 mg LRG, 3.5 mg R, 4.0 mg N, 6.0 mg M, 11.3 mg Y, 55.2 mg UMP, and 90.1 mg Fru were dissolved in 10 mL ultrapure water to achieve final concentrations of 0.5, 2, 3, 4, 5, 15, and 50 mM, respectively. With an MspA-FPBA, 10 μL of the above-described solution was added to the cis chamber. A transmembrane voltage of +160 mV was continually applied during the measurements. Events were automatically predicted by the previously trained model and labeled accordingly. (c) The scatter plot of ΔI versus SD for the M, N, Y, R, UMP, Fru, and LRG events obtained from a-b. The scatter plot was generated using a 30 min continually recorded trace. 1072 successive events were employed to generate the statistics. (d) Comparison of the concentrations of Fru, UMP, Y, M, N, R, and LRG reported by nanopore analysis with the true values. The values measured in this work were determined and calibrated as described in Methods. For the quantitative analysis, three independent measurements (N = 3) were conducted to generate statistics. Data are presented as mean ± standard deviations, with error bars representing standard deviation values.

Table RL 4.1 Quantitative analysis of amino acids, NMPs, saccharides, and peptides in the standard sample.

Analytes	True concentration value (mM)	Nanopore measurement value (mM)
LRG	0.5	0.66 ± 0.27
Arg	2	1.73 ± 0.41

Asn 3 2.43 ± 0.26
Met 4 3.37 ± 0.22
Tyr 5 4.45 ± 0.31
UMP 15 12.25 ± 1.33
Fru 50 45.12 ± 1.24

To this end, we confirm that it is feasible to perform quantitative analysis of individual species in specific mixtures. This demonstration provides independent experimental evidence supporting the species identities predicted by the model in mixed sensing. In the revised manuscript, all relevant results (Supplementary Tables 11-12 and Supplementary Figs. 30-36) have been added to complement the discussions.

We appreciate Reviewer 4's insightful comments on protein sequencing. We concur with the reviewer's view that the integration of machine learning with protease digestion experiments represents a promising approach for determining complete protein sequences in future research.

Finally, acknowledging your inspiring and insightful suggestion, the quality of this manuscript has been further improved, which is highly appreciated. Some additional discussions were also added to the revised manuscript to fully address all your raised issues. For the ease of your reference, they are also pasted below as well.

"During the detection of amino acids, NMPs, saccharides, and peptides, within a certain analyte concentration range, the count of nanopore events generated per min by a specific species (F) is linearly correlated with the species concentration, forming the basis for quantitative analysis (Supplementary Tables 11-12 and Supplementary Figs. 30-36). To demonstrate that, four amino acids (Met, Asn, Tyr, and Arg), UMP, Fru, and the tripeptide LRG were mixed at known concentrations to generate a mixture (Methods). The mixture was subsequently subjected to nanopore analysis (Supplementary Fig. 37). According to the quantitative results, the nanopore results were generally consistent with the true values (Methods and Supplementary Fig. 37). These results not only demonstrate its quantitative capability for mixed samples, but also validate the machine learning model's capability for mixture analysis."

Review Comments:

The machine-learning methodology is generally appropriate in the context of this work and contributes to the strength of the study. Our specific comments are below. The use of bagged trees ensemble achieves a near-99% accuracy under cross-validation, which is plausible given the distinct blockage currents for most of the molecules. The paper demonstrates an effective use of classical ML approaches combined with well-designed features to distinguish the signals from different molecules. However, our own reproduction of the model using the provided code files, suggests that a smaller subset of features {'Mean', 'SD', 'toff', 'LocalSD_Ratio'} is capable of achieving performance similar to the 9-feature model. Furthermore, fitting the provided data to Random Forest models shows a capability to achieve similar scores using pretty low n_estimators (3-5), and max_depth (10-20) parameters. Such near-maximal accuracy with such a lightweight model strongly suggests that the underlying feature space is highly separable and that several of the nine engineered features may be redundant. A brief analysis of the minimal feature subsets would significantly strengthen claim of the multi-functional nanopore producing a consistently informative and discriminative signals.

Response:

We appreciate Reviewer 4's recognition of the model's near-perfect 99% cross-validation accuracy. We are encouraged by the acknowledgment of our methodological rigor in selecting appropriate classical machine learning approaches and identifying discriminative event features. The insights provided regarding minimal feature subset analysis are especially insightful, as they demonstrate the technique's capacity to consistently generate informative and discriminative nanopore signals. These constructive comments enhance our understanding of the method's potential and its practical applications.

In response to the reviewer's suggestion, we trained the model using a minimal feature subset (Mean, SD, toff, and LocalSD Ratio). As noted by the reviewer, this 4-feature model demonstrated performance comparable to the original 9-feature model (Table RL 4.2), suggesting the feasibility of a lightweight model configuration. However, this performance equivalence was specifically observed in the context of distinguishing among 29 analytes, comprising 21 amino acids, 4 nucleotide monophosphates, and 4 saccharides.

Table RL 4.2 Accuracy of distinguishing 29 analytes using the 4-feature model and the 9-feature model, respectively.

Bagged Trees Model	Validation Accuracy	Test Accuracy
4-feature model	0.9882	0.9843
9-feature model	0.9887	0.9898

When the analytical scope is further expanded to include post-translationally modified amino acids, epigenetic nucleotide monophosphates, and peptide sequences (increasing the total analyte count to 40), the performance divergence became more pronounced (Table RL 4.3). Specifically, the 4-feature model exhibited a 0.42% decrease in validation accuracy and a 0.92% decrease in test accuracy compared to the 9-feature model. This divergence highlights the role of comprehensive feature sets in maintaining model robustness for complex chemical classification tasks involving diverse analyte groups.

Given the comparable computational costs between the 4-feature and 9-feature models during training and the advantages of the 9-feature model in the analysis of a more expanded analytical scope, we opted to retain the 9-feature configuration throughout our analysis. This choice offers practical advantages for future model expansion, particularly in accommodating additional analytes to the current framework in prospects. Though not demonstrated in this paper, it is worth noting that the underlying nanopore detection principles could theoretically be extended to other molecular classes, including D-amino acids, nucleotide sugars, oligosaccharides, and oligonucleotides.

We sincerely appreciate the reviewer's thoughtful suggestion regarding model simplification and your insightful contribution to exploring lightweight alternatives. In response, we have included additional supporting information results in the revised manuscript that systematically evaluate the 4-feature model's potential as a streamlined option, along with relevant discussions of its performance trade-offs.

Table RL 4.3 Accuracy of distinguishing 40 analytes using the 4-feature model and the 9-feature model, respectively.

Bagged Trees Model	Validation Accuracy	Test Accuracy
4-feature model	0.9828	0.9585
9-feature model	0.9870	0.9677

Eventually, we again thank reviewer 4 for raising these extremely insightful and constructive comments and valuable suggestions. In the revised manuscript, we have added a brief discussion of the minimal feature subsets. For the ease of your reference, these added texts are also pasted below.

“The 4-feature model trained on the minimal feature subset (ΔI, SD, LocalSD Ratio, and toff) exhibited performance comparable to the original 9-feature model (Supplementary Table 9). This highlights the potential for a lightweight model configuration. However, this equivalence in performance was observed specifically in the context of differentiating 29 analytes (Fig. 2a). When the analytical scope was expanded to 40 analytes, including amino acids with post-translational modifications (PTMs), epigenetic nucleoside monophosphates (NMPs), and peptides (Extended Data Fig. 6b), the performance gap became more evident (Supplementary Table 10). To maintain consistency and in anticipation of future broadening of the analytical scope, we retained the 9-feature model throughout the manuscript. While not demonstrated here, the underlying principle is also theoretically applicable to other molecular classes, such as D-amino acids, nucleotide sugars, oligosaccharides, and oligonucleotides.”

Review Comments:

The code files provided only include the code and data for the AAs_NMPs_Sacs_classifier (used for Fig. 3 and Supplementary Table 6). The code and data used for the few-shot learning method described in Figure 5, as well as that which is used for identification of the PTM modified amino acids, epigenetic NMPs and peptides (and for Extended Data Figure 5) are missing from the source file. This makes it difficult to reproduce the study and verify the performance of the ML models. We recommend providing this to the reviewers and also to the readers eventually.

Response:

As requested by reviewer 4, we have now provided the codes for the few-shot learning and the identification of amino acids with PTMs, epigenetic NMPs, and peptides.

Review Comments:

Lines 281-283: The authors note that the model used for the out-of-database classification was trained on the same dataset described in Fig. 3, which was generated exclusively from single-analyte experiments. While the few-shot generalization to new analytes is encouraging, extending a model trained solely on isolated analytes to scenarios involving multi-analyte mixtures requires additional justification. In multi-species experiments, event characteristics may be altered by competitive interactions, concentration imbalances, or subtle changes in translocation dynamics. Demonstrating that the event distributions from single-analyte experiments remain valid in these more complex conditions, would considerably strengthen the manuscript.

Response:

We deeply appreciate Reviewer 4's insightful comments and suggestions regarding the challenges of applying machine learning models trained on single-analyte datasets to multi-analyte mixtures. We fully agree with the reviewer's perspective that analyte interactions in multi-analyte mixtures can introduce uncertainties in predictive outcomes. To further address this concern, we conducted comprehensive mixture analyses using random combinations of analytes (Figure RL 4.3). The resulting experimental data demonstrate strong validity and consistency with the results obtained from individual analyte measurements (Figure RL 4.4). These findings affirm the robustness of our approach under diverse mixture conditions. It also suggest that the molecular interaction

Figure RL 4.3 Nanopore sensing of a multi-analyte mixture containing Gly (G), Met (M), Gln (Q), Arg (R), UMP, GMP, Fru, the dipeptide IK, and the tripeptide LRG. (a) Continuous current traces acquired from the multi-analyte mixture. The mixture was added to the cis chamber to achieve final concentrations of G (85 μM), M (80 μM), Q (100 μM), R (60 μM), UMP (300 μM), GMP (400 μM), Fru (1 mM), IK (15 μM), and LRG (10 μM). A transmembrane voltage of +160 mV was continually applied during the measurements. Events were automatically predicted by the previously trained bagged trees model and labeled accordingly. (b) The scatter plot of ΔI versus SD for the G, M, Q, R, UMP, GMP, Fru, IK, and LRG events obtained from a. The scatter plot was generated using a 30 min continually recorded trace. 947 successive events were employed to generate the statistics.

between the selected analytes are negligible in this assay and won't interfere with machine learning prediction.

Briefly, Gly (G), Met (M), Gln (Q), Arg (R), UMP, GMP, Fru, the dipeptide IK, and the tripeptide LRG were selected to prepare a defined mixture. The mixture was introduced into the cis chamber to achieve final concentrations of 85 μM G, 80 μM M, 100 μM Q, 60 μM R, 300 μM UMP, 400 μM GMP, 1 mM Fru, 15 μM IK, and 10 μM LRG.

Figure RL 4.4 Comparison of nanopore events obtained from single-analyte and mixture experiments. (a) Visual comparison of representative events of analytes (G, M, Q, R, UMP, GMP, Fru, IK, and LRG) obtained from single-analyte experiments and from the mixture experiment presented in Figure RL 4.3. Single-analyte events and mixed-analyte events are indicated in gray and color, respectively. No significant differences in the event characteristics of each analyte were observed. (b) Overlap of event distributions between events acquired with single-analyte and mixed analytes. Gray points represent events from single-analyte experiments, with red ellipses indicating the 99% confidence regions. Colored points correspond to events recorded from mixture experiments, which largely fall within the same regions. To more clearly show the details, the events inside the blue box are further zoomed-in and shown on the right.

Subsequently, mixed-analyte events (color-coded) were compared with corresponding single-analyte events (gray-coded). Visually, no significant differences in event characteristics were observed between mixed-analyte and single-analyte events (Figure RL 4.4a). To further analyze the data, we constructed 2D scatter plots of ΔI versus SD and compared the event distributions from single-analyte and mixed-analyte experiments (Figure RL 4.4b). Single-analyte events are represented by gray scatter points, with 99% confidence ellipses derived from their distributions. Mixed-analyte events from the mixture experiments are overlaid as colored points. As shown in Figure RL 4.4b, the mixed-analyte events clearly overlap with single-analyte distributions and reside within the corresponding confidence regions, demonstrating that the event distributions obtained under single-analyte conditions are consistent with those observed in complex multi-analyte

environments.

In addition, based on our previous quantitative analysis of a standard mixture sample (Figure RL 4.2 and Table RL 4.1), in which the component concentrations varied significantly from 0.5 to 50 mM. Under such concentration-imbalanced conditions, our method can still clearly identify each component in the mixture and achieve accurate quantification. This indicates that event characteristics in multi-species experiments are not affected by concentration imbalance.

Finally, we extend our sincere appreciation to Reviewer 4 for raising this insightful discussion, which led to significant additions in the revised manuscript. Specifically, we have incorporated enhanced discussions on event characterization under mixed sensing and quantification capabilities under concentration-imbalanced conditions. These revisions have not only addressed the raised discussions but also substantially improved the manuscript's overall quality. For your convenience, the relevant additions have been highlighted and are pasted below:

“To demonstrate the identification performance of the model trained on the single-analyte dataset for multi-analyte mixtures, we performed nanopore sensing on a defined multi-analyte mixture (Supplementary Fig. 38) and compared the produced events with those from corresponding single-analyte measurements (Supplementary Fig. 39). Quantitative analysis through 2D scatter plots of ΔI vs. SD demonstrated that the event signatures in multi-analyte mixtures are identical to that collected during single-analyte measurements, with all events falling within the 99% confidence ellipses derived from the training data (Supplementary Fig. 39), confirming the model's consistency across different sample matrices. Notably, our quantitative evaluation of a reference mixture featuring a 100-fold concentration gradient (0.5-50 mM) (Supplementary Fig. 37) revealed no loss in analytical accuracy, demonstrating the method's robustness to concentration imbalances in multi-species environments.”
Reviewer #4 (Remarks on code availability):

Review Comments:

See main review, some of the code was not made available and needs to be made available. We were able to run the code provided and provide some comments in the main review.

Response:

As requested by reviewer 4, we have now provided the codes for the few-shot learning and the identification of amino acids with PTMs, epigenetic NMPs, and peptides, corresponding to the results as described in the session of “Code statement”. We thank reviewer 4 for providing valuable suggestions. The relevant description is also provided below.

“The custom machine learning codes and training data for the identification of amino acids, NMPs, saccharides, and peptides, as well as for few-shot learning, are available in the attachment as “Unified Classifier”.”
Reviewer #5:

Review Comments:

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Biotechnology initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Response:

We sincerely appreciate reviewer 5's insightful suggestions and valuable time. The questions raised by you and your co-reviewer are greatly appreciated and have all been well addressed in a point-by-point manner. At anytime, please never hesitate to let us know if there is any remaining unclear issues, we will be more than happy to clarify it.
Reviewer #6:

Review Comments:

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Biotechnology initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Response:

We sincerely appreciate reviewer 6's insightful suggestions and valuable time. The questions raised by you and your co-reviewer are greatly appreciated and have all been well addressed in a point-by-point manner. At anytime, please never hesitate to let us know if there is any remaining unclear issues, we will be more than happy to clarify it.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have included additional experiments, data, and discussion, which address our major concerns. The addition of the distinguishing isomers figure and the LOD tables is very powerful. We ask that the authors clarify that the LOD concentration definition derives from one reconstituted nanopore. It would be more interesting if determined for multiple reconstituted nanopores and included in the standard deviation. However, this is not essential. Also, the addition of the non-glycosylated peptide analysis and all mass spec measurements were critical.

A few small concerns should be addressed. Listed below

The authors need to more clearly state that longer peptides create too much steric hindrance and are not classified by MspA-FPBA.
Please add a description in the figure legends (figures 2b-d, 4 b-c, and 5b-c) to clarify the difference between each represented trace.
In the extended data figure 4f, a scale for the current pA would make it more informative.
In the extended data figure 7 b, the top trace is missing an axis label.

Other than these small things, this paper is ready on my end to be published in Nature Biotech.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Biotechnology initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
After going through the revised manuscript and the responses to the critique, I believe this manuscript is ready for publication as is. I think the authors have done a spectacular job of addressing the potential weaknesses of the method, by thoroughly addressing each point with a thoughtful response. One drawback remains the low sensitivity to various analytes, although as the authors mention, scaling up the number of nanopores (n pores) would address that by reducing input requirements n-fold (approximately). Another potential drawback is unknown analyte interference, which can appear to be coming from a known substrate but really comes from an unchartered analyte. However, it is important to realize that nanopores are not mass spectrometers, and one has to identify and calibrate all interfering agents in order to obtain meaningful results. Using internal standards and other calibration techniques one can overcome these limitations to a meaningful extent.

(Remarks on code availability)
I have not reviewed the code.

Reviewer #4

(Remarks to the Author)
We have closely reviewed all the revisions as well as the code submitted. We appreciate the efforts made by the authors to respond to comments and are satisfied with the changes made. We have no further comments and feel like this would be a good contribution to the journal.

(Remarks on code availability)
The code works well and is reproducible and accesible.

Reviewer #5

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Biotechnology initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #6

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Biotechnology initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)