PNAS nexus

PNAS Nexus, 2024, **3**, pgae350
https://doi.org/10.1093/pnasnexus/pgae350
Advance access publication 24 September 2024
Research Report

A breath-based in vitro diagnostic assay for the detection of lower respiratory tract infections

Dapeng Chen[a,*], Marek A. Mirski[b], Shuo Chen[c,d], Wayne A. Bryden[a], Michael McLoughlin[a], Kiana M. Kiser[a], Emily R. Caton[a], Caroline R. Haddaway[a], Maximilian S. Cetta[a] and Yezhi Pan[d]

[a]Zeteo Tech, Inc., Sykesville, MD 21784, USA
[b]Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[c]Division of Biostatistics and Bioinformatics, Department of Epidemiology and Public Health, School of Medicine, University of Maryland, Baltimore, MD 21205, USA
[d]Department of Psychiatry, Maryland Psychiatric Research Center, School of Medicine, University of Maryland, Baltimore, MD 21201, USA
*To whom correspondence should be addressed: Email: dapeng.chen@zeteotech.com
Edited By David Brenner

Abstract

An accurate diagnosis is critical to reducing mortality in people with lower respiratory tract infections (LRTIs). Current microbiological culture is time-consuming, and nucleic acid amplification-based molecular technologies cannot distinguish between colonization and infection. Previously, we described developing a sampling system for effectively capturing biomolecules from human breath. We identified a new class of proteoform markers of protease activation, termed proteolytic products of infection, for detecting LRTIs in people with mechanical ventilation. Here, we further developed an in vitro assay by designing a specific substrate sensor for human neutrophil elastase (HNE) to detect LRTIs in breath samples. In the proof-of-concept study, we then applied this in vitro assay to breath samples collected from intubated patients and healthy volunteers. The findings revealed that the LRTI group demonstrated a significant mean differential, showing a 9.8-fold elevation in measured HNE activity compared with the non-LRTI group and a 9.2-fold compared with healthy volunteers. The in vitro assay's diagnostic potential was assessed by constructing a receiver operating characteristic curve, resulting in an area under the curve of 0.987. Using an optimal threshold for HNE at 0.2 pM, the sensitivity was determined to be 1.0 and the specificity to be 0.867. Further correlation analysis revealed a strong positive relationship between the measured HNE activity and the protein concentration in the breath samples. Our results demonstrate that this breath-based in vitro assay provides high diagnostic performance for LRTIs, suggesting that the technology may be useful in the near term for the accurate diagnosis of LRTIs.

Keywords: human breath, noninvasive diagnostics, mass spectrometry, lower respiratory tract infection

Significance Statement

Using human breath as a noninvasive approach to diagnosing diseases has gained broad attention over the years, but the absence of effective breath collection technologies and practical assays has inherently hindered progress. To fill the gap, we developed an in vitro assay based on breath collection and a sensitive protease-substrate sensor detecting dysregulated protease activity, a hallmark of bacterial infections, for detecting lower respiratory tract infections (LRTIs). We demonstrated the sensitivity of this assay and further evaluated it using clinical breath samples from intubated patients with LRTIs. Our proof-of-concept study demonstrated the potential for LRTI diagnosis using this method and supported developing in vitro diagnostic tools that use human breath as a noninvasive approach for disease detection and diagnosis.

Introduction

Lower respiratory tract infections (LRTIs), including bronchitis, pneumonia, and bronchiolitis, have been reported as the fifth-leading cause of mortality globally (1). According to the 2015 Global Burden of Disease Study, these infections accounted for an estimated 2.74 million fatalities (1). The mortality figures were substantially higher when factoring in tuberculosis (TB) and the COVID-19 pandemic. The management of LRTIs relies on accurate diagnosis followed by antimicrobial therapy (2). Thus, timely diagnosis is the cornerstone of effective management of LRTIs. In practice, the gold standard for LRTI diagnosis is quantitative microbiological culture of respiratory specimens, which enables the isolation and identification of potential bacterial pathogens (2). However, culture-based methods are

Competing Interest: D.C., K.M.K, C.R.H., E.R.C., M.S.C., W.A.B., and M.M. have competing interests. D.C., K.M.K., C.R.H., M.S.C., and E.R.C. are employed by Zeteo Tech, Inc. W.A.B. and M.M. serve as chief executive officer and chief technology officer for Zeteo Tech, Inc., respectively. M.A.M., Y.P., and S.C. have no competing interests. An unpublished US Provisional Patent Application assigned to Zeteo Tech, Inc. was applied based on this research.
Received: March 29, 2024. **Accepted:** August 7, 2024
OXFORD UNIVERSITY PRESS
© The Author(s) 2024. Published by Oxford University Press on behalf of National Academy of Sciences. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

labor-intensive and time-consuming, which causes delays in effective antimicrobial therapy.

Molecular techniques targeting pathogen markers have emerged as alternatives to traditional culture methods. In particular, nucleic acid amplification tests with multiplexed capacity have been used to improve LRTI diagnosis, such as CE marked Unyvero HPN panel by Curetis and FDA-cleared FilmArray PN Panel by bioMérieux (3, 4). Molecular assays offer speed and high sensitivity. However, their dependence on pathogen materials presents a challenge in distinguishing between infection and colonization, wherein identical test outcomes may suggest disparate clinical trajectories, given the extensively reported colonization of the respiratory tract (5). Moreover, pathogen marker-based molecular assays cannot assess bacterial viability, further limiting its application in the accurate diagnosis of LRTIs. Therefore, host response factors have been investigated to enhance the diagnostic accuracy and the clinical management of patients with LRTIs (6–11). Among those markers, proteases, particularly those derived from neutrophils, have been extensively investigated due to the robust scientific evidence showing their dysregulation during LRTIs (7–10).

Particles expelled in human exhaled breath, originating from the lower respiratory tract via the airway reopening mechanism, contain nonvolatile organic compounds (NOCs), including peptides, proteins, and RNA (12–14). When interfaced with sophisticated analytical methodologies, these NOCs hold promise for disease diagnosis (15–17). However, exploring the diagnostic potential of NOCs has historically been constrained by the need for effective collection technologies capable of capturing submicron particles that contain these organic compounds in human breath. To overcome the challenge, we developed an advanced breath collection system based on chemical affinity, BreathBiomics, to explore using human breath as a diagnostic strategy (14). In particular, we identified a new class of human response markers, termed proteolytic products of infection (PPI), and demonstrated their link with LRTIs (17). These host response factors, resulting from protease dysregulation—a hallmark in LRTIs—hold promise for detecting LRTIs in intubated patients (17).

Here, we extended the exploration of PPI application by engineering a substrate sensor targeting human neutrophil elastase (HNE), thereby developing a breath-based in vitro assay to detect HNE activity within human breath. We applied this newly developed methodology to intubated patients with mechanical ventilation in intensive care units (ICUs), aiming to detect LRTIs. Our results indicate that this breath-based in vitro assay could be used as a noninvasive approach to detect LRTIs within a critical care setting.

Materials and methods

HNE substrate sensor

PEG36-Nle(O-Bzl)-Met(O)2-Oic-Abu-ACC was designed in house and synthesized by CPC Scientific Inc (San Jose, CA, Fig. S1). Amino-PEG36-acid (CAS: 196936-04-6) and 2-(7-amino-2-oxochromen-4-yl)acetamide (CAS:296236-23-2) were used for the synthesis. The synthesized sensor was resuspended in high-performance liquid chromatography (HPLC)-grade water with a concentration of 200 μM.

HNE substrate sensor incubation with rmNE and rhNE

Recombinant mouse neutrophil elastase (rmNE) (Catalog # 4517-SE), recombinant human neutrophil elastase (rhNE) (Catalog # 9167-SE), and recombinant mouse active cathepsin C/DPPI (rmCatC, Catalog # 2336-CY) were purchased from R&D Systems (Minneapolis, MN). Native human cathepsin G (CTSG) protein (Catalog # ab91122) was purchased from Abcam. Human neutrophil PR3 (Product # 16-14-161820) was purchased from Athens Research and Technology, Inc; 0.75 μM of rmNE or rhNE was activated by 1 mM rmCatC in 10 μL of activation buffer (50 mM MES, 50 mM NaCl, pH 5.5) for 2 h at 37 °C. After activation, 15 nM of rmNE was incubated with 100 μM in 100 μL of assay buffer [50 mM Tris, 1 M NaCl, 0.05% (w/v) Brij-35, pH 7.5] at different incubation conditions, including different incubation time and temperature. In the LOD test conducted for the study, nine dilutions of rhNE were prepared, ranging from 15 nM to 0.04 pM. For this series, rhNE was diluted in the assay buffer. All experiments were repeated at least three times to ensure reproducibility. The reproducibility was demonstrated by standard deviation.

High-Resolution mass spectrometric characterization of the sensor and the cleavage product

Samples were introduced to an LTQ orbitrap mass spectrometer (Thermo Fisher Scientific) with ESI capacity via direct infusion as a flow rate of 6 μL/min. Precursor profiles were acquired in the positive ion mode with the resolution of 60,000. Ion fragmentation profiles were acquired with 35% collision-induced dissociation (CID) energy with an isolation window of 1.5 *m/z*. The raw mass spectrometry files were processed, and deconvoluted profiles were acquired using FreeStyle Software (Thermo Fisher Scientific). Deconvoluted profiles and ion fragmentation patterns were interpreted manually in house.

MALDI-TOF mass spectrometric characterization of the sensor and the cleavage product

For MALDI-TOF MS, 1 μL of the sample was added onto a Bruker MALDI plate and dried. Then, 1 μL of α-cyano-4-hydroxycinnamic acid MALDI matrix (9 mg/mL in 70% acetonitrile) was added to the sample and dried. MALDI-TOF MS profiles were acquired from Bruker Daltonics microflex LRF mass spectrometer (Billerica, MA) in the positive linear mode. Mass spectra were obtained from 500 profiles in the mass range of 1,500 to 3,000 *m/z*. The ion intensity values of the mass peaks of interest were extracted from the raw files generated directly from the operation software on the Bruker instrument.

Bottom-up proteomics

Ten microliters of breath sample was used for the analysis. For the experiment, 50 μL of 50 mM ammonia bicarbonate (pH 8.5) was added to each sample. Protein reduction was achieved by adding 5 mM of dithiothreitol (DTT), followed by incubation at 37 °C for 30 min. Subsequently, protein alkylation was conducted by adding iodoacetamide to achieve a final concentration of 15 mM, with the mixture then incubated at room temperature for 1 h. Sequencing Grade Trypsin (Promega Corporation) was used for overnight protein digestion. Postdigestion, the peptides were desalted using in-house packed C18-packed tips. The peptide samples were then suspended in 20 μL of 0.1% formic acid for subsequent mass spectrometry analysis. The samples were introduced to the mass spectrometer using an EASY-nLC 1000 system coupled with a Q-Exactive HF Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific). In the tandem mass spectrometry analysis, peptides were loaded onto an Acclaim PepMap 100 C18 trap column (0.2 mm × 20 mm, Thermo Fisher Scientific) at a flow rate of 5 μL/min and separated on an EASY-Spray HPLC Column (75 μm × 150 mm, Thermo Fisher Scientific). The HPLC gradient is composed of 5–55% of the mobile

phase (75% acetonitrile and 0.1% formic acid) at a 300 nL/min flow rate over 60 min. Mass spectrometry data collection was performed in the data-dependent acquisition (DDA) mode, with key parameters including a precursor scanning resolution set to 60,000 and a product ion scanning resolution of 15,000. Product ion fragmentation was acquired using high-energy CID at 35% total energy. The bottom-up proteomics raw data files were analyzed using MaxQuant-Andromeda software (maxquant.org) against the human protein database (Homo sapiens, uniprot.org) with the reviewed protein entries of 26,816. Ion intensity of identified protein entries was provided by the database searching software and used for protein abundance analysis. When a protein was not detected in the analysis, the ion intensity for that specific protein was calculated by assigning an ion intensity value two orders of magnitude lower than the default precursor ion intensity, which is 1E4 (10,000).

Study participants

The breath samples were collected from the Neurosciences Critical Care (NCCU) of Anesthesiology and Critical Care Medicine at the Johns Hopkins Hospital (JHH-NCCU). The institutional review boards (IRBs) at The Johns Hopkins University School of Medicine (Application number: IRB002494495) approved this study, and all experiments included in this study were conducted based on the approved protocols. Study participants were provided with written informed consent. Per our IRB submission and approval, patients meeting eligibility criteria (or legally authorized representatives consenting for the patient) were approached by the study team for consent.

Patient data included age, gender, race, ethnicity, primary diagnosis, medication regimen, the timing of sample collection, microorganism identification information, results of white blood cell tests, body temperature, fraction of inspired oxygen (FiO_2) test results, and pulmonary radiography findings. The process of microorganism identification was performed in the clinical laboratory at the JHH and subsequently interpreted by the ICU clinicians. The physicians at the JHH diagnosed positive respiratory tract infection, utilizing established clinical criteria. Confirmation of these diagnoses was achieved when respiratory tract samples, including sputum, endotracheal tube samples, or BAL, yielded positive cultures in the clinical laboratory at the JHH.

Healthy volunteers were briefed on the experimental protocol and provided their informed consent before conducting the study. The data derived from this research was processed in a way that excluded personally identifiable information and was not intended for regulatory approval processes.

Exhaled breath collection from intubated patients

The BreathBiomics system was interfaced with the mechanical ventilator using a T fitting attached to the ventilator's exhaust tubing to collect breath samples (Fig. 1A). This setup allowed the exhaled breath to be directed into the collection column through a pump system, which comprises a CO_2 sensor (Gas Sensing Solutions Ltd, Cumbernauld, UK) and a mini diaphragm pump

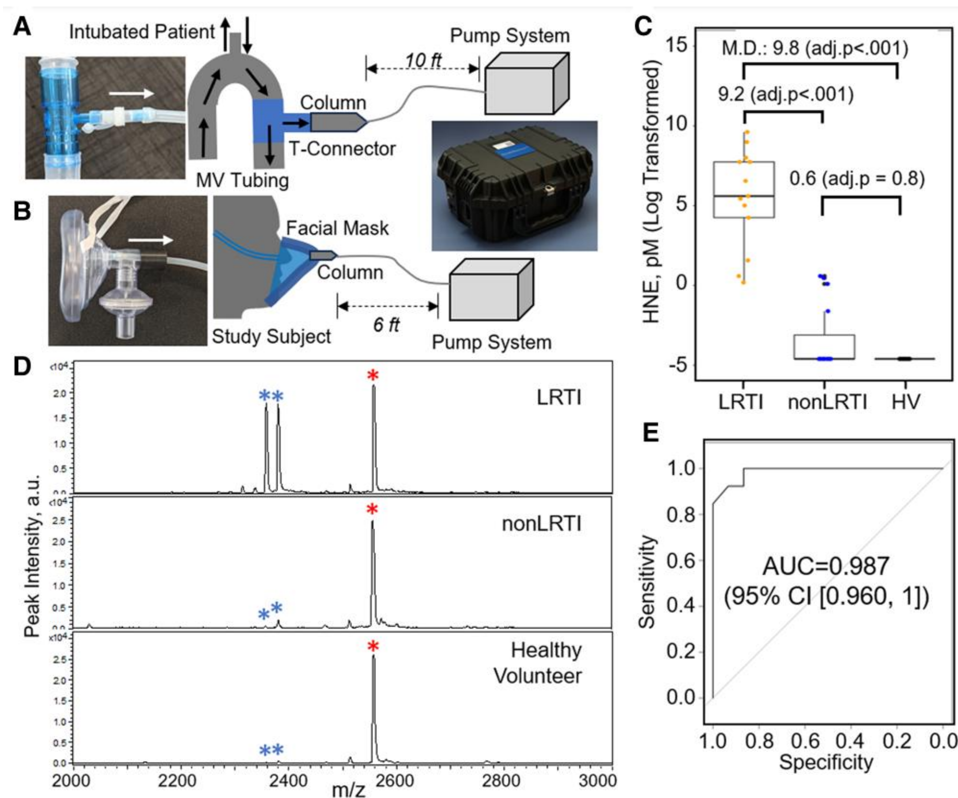

Fig. 1. Evaluation of the diagnostic potential of the in vitro assay using breath samples collected by BreathBiomics. *A*) Schematic of breath collection from intubated patients on mechanical ventilation in ICUs and the pump system. *B*) Schematic of breath collection from healthy volunteers using a facial mask. *C*) Boxplot display of the distribution of natural-log-transformed HNE measurements, mean differences, and adjusted *P*-values by group: LRTI ($n = 13$), non-LRTI ($n = 15$), and healthy volunteers (HV; $n = 19$) *D*) Representative MALDI-TOF spectra of breath samples collected from different groups. *E*) ROC curve analysis for the in vitro assay performance.

(Parker Hannifin Corporation, Cleveland, OH). The flow rate of the pump was calibrated to 0.5 L/min so as not to interfere with calculated exhaled ventilation. The CO_2 sensor recorded the levels of exhaled CO_2 in the ventilator's exhaust tubing, validating the closed system and continuous sampling. Upon completion of the collection process, the BreathBiomics collection column was removed from the ventilator, decontaminated in the clinical laboratory at the JHH, and sent to Zeteo Tech's laboratory for further sample processing. At Zeteo Tech's laboratory, 300 µL of 70% acetonitrile (volume/volume in water) was added into the column and pushed using a syringe to extract and collect the biomolecule materials. Acetonitrile was then removed through overnight lyophilization. After that, 50 µL of HPLC-grade water was added to the residue. These prepared samples were stored in the laboratory's refrigerator, pending further analysis.

Exhaled breath collection from healthy volunteers

The BreathBiomics column was integrated into a CPR facial mask through its front port (Fig. 1B). This column was connected to a pump system identical to the one used for intubated patients, with the difference being adjusting the flow rate to 2.5 L/min. Throughout the breath collection experiment, the study subjects were permitted to breathe naturally, without any imposed restrictions, and each collection session was set to 10 min. After completing these sessions, the columns were treated to the same processing protocol as previously described in intubated patients.

In vitro assay using clinical breath samples

In the in vitro assay protocol, 5 µL of resuspended breath sample was mixed with the sensor mixture that included 15 µL of a 200 mM substrate sensor prepared in the assay buffer. The sample was then incubated for 24 h at 37 °C. After incubation, 1 µL of the sample was used for MALDI-TOF MS.

Data analysis and statistics

We conducted an analysis of covariance (ANCOVA) to assess the disparities in HNE levels across the three groups (LRTI, non-LRTI, and healthy volunteers), with age and sex as covariates. Subsequent to the initial ANCOVA, we conducted pairwise comparisons to examine specific group differences. For this purpose, we applied Tukey's post hoc test, which generated adjusted *P*-values using Tukey's HSD (honestly significant difference) method. All HNE levels were natural-log-transformed for normalization, and prior to log transformation, a constant of 0.01 was added to all zero values to prevent undefined logarithmic operations. We used a logistic regression model to examine the relationship between the LRTI status and the HNE level. To assess the performance of the logistic regression model and determine its predictive accuracy, we calculated the AUC–ROC. Then we employed Youden's Index to identify the optimal classification threshold that maximized the balance between sensitivity and specificity. All analysis was performed in R.

Results

Overview of the breath-based in vitro assay and design of the HNE substrate sensor

This breath-based in vitro assay for detecting LRTIs has four main steps, as illustrated in Fig. 2A and Fig. S1. Prior studies have demonstrated that proteases were dysregulated during bacterial infection, leading to their identification in lower respiratory tract specimens, including particles emitted from the breath (8–10, 17). BreathBiomics efficiently captures these proteases, as represented by the the small dots in Fig. 2A. Upon capture, these proteases are incubated with a customized substrate sensor to HNE in vitro. After incubation, the HNE cleavage product and intact sensor are characterized with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), a

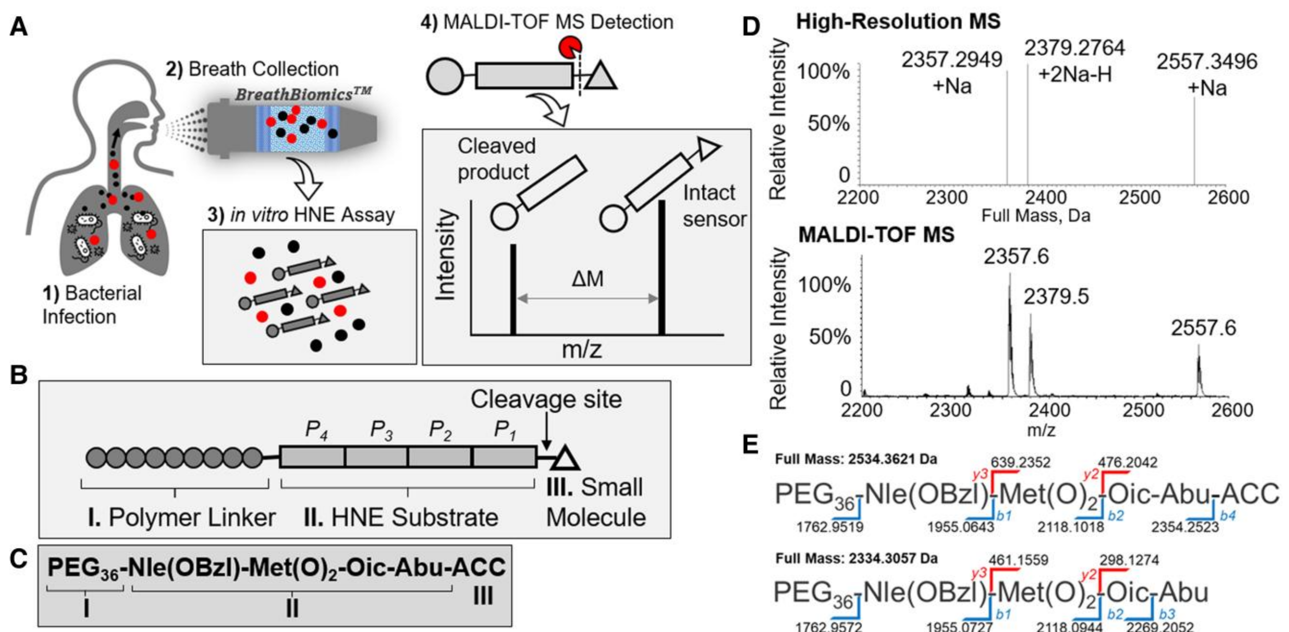

Fig. 2. Bread-based in vitro assay overview and HNE substrate design. *A*) Overview of the breath-based in vitro assay to detect LRTIs. *B*) The concept of the HNE substrate sensor designed in this study. *C*) The structure of the HNE substrate sensor used in this study. *D*) Representative high-resolution and MALDI-TOF mass spectra of the HNE substrate sensor and its cleavage product. *E*) Ion fragmentation maps of HNE substrate sensor and its cleavage product characterized by DI-ESI-MS/MS.

rapid spectrometric technique (18). Based on the mass intensity values extracted from MALDI-TOF MS and the ratio between the intact sensor and its cleavage product, we proposed that the in vitro assay offers a quantitative assessment of HNE activity.

Our customized HNE substrate sensor, PEG36-Nle(O-Bzl)-Met(O)2-Oic-Abu-ACC, is composed of a polymer linker, an HNE substrate, and a cleavable small molecule (Fig. 2B). We used a substrate made of unnatural amino acids, Nle(O-Bzl)-Met(O)2-Oic-Abu, which is reported to have a higher sensitivity and specificity than conventional substrates to HNE (Fig. 2C) (19). A subsequent covalent modification at the C-terminus with 7-amino-4-carbamoylmethylcoumarin (ACC) facilitates its removal by HNE. To tune the sensor's mass within the optimal detection window of 1,500 to 20,000 *m/z* for MALDI-TOF MS and to avoid interference from detergent signals generated from the buffer system such as Brij-35 detergent, the substrate's N-terminus is conjugated covalently to amino-PEG36-acid (polyethylene glycol, PEG).

High-resolution mass spectrometry and MALDI-TOF MS were used to characterize the sensor and its cleavage product after its interaction with a rmNE (Fig. 2D). The sensor's mass peak of 2,557.6 *m/z* was identified as sodium adduct. The cleavage product exhibited two mass peaks at 2,357.6 and 2,379.5 *m/z*, indicative of the same molecule but with varied sodium adducts. We further conducted our established method to use direction-infusion electrospray ionization (ESI) tandem mass spectrometry (DI-ESI-MS/MS) to characterize the ion fragmentations of both the sensor and its cleavage product (Fig. 2E and Fig. S1) (20). The results showed an ACC removal in the rmNE's presence, a deduction substantiated by the mass difference of 200.0564 Da, the molecular structure of 2-(7-amino-2-oxochromen-4-yl) acetamide (C11H8N2O2, Fig. S1). Moreover, b ions (bottom symbols) and y ions (top symbols) of the unnatural amino acids of the HNE substrate were observed in both the sensor and the cleavage products. Collectively, we designed an HNE substrate sensor for the in vitro assay based on an established HNE substrate. High-resolution mass spectrometric analyses confirmed its structure, the cleavage site, and the cleavage product.

Evaluation of the in vitro assay performance using rmNE and recombinant human neutrophil elastase

We investigated the feasibility of our in vitro assay using rmNE with two primary objectives: first, to evaluate the performance of our customized HNE substrate sensor, and second, to construct a calibration curve for quantitation. Our results showed that as early as 10 min after incubation with 15 μM of rmNE, cleavage product mass peaks were detected by MALDI-TOF MS (Fig. 3A). Furthermore, after an incubation period of 3 h, the intact sensor peak was no longer present, suggesting a complete enzymatic reaction, as shown by the blue and red regions in Fig. 3A. Further quantification of the reaction kinetics was achieved by calculating the ratio of the cleavage product to the intact sensor, extracted from the mass peak intensities shown in Fig. 3B. Notably, the two cleavage product mass peaks, 2,357 and 2,379 *m/z*, exhibited a strong correlation (Fig. 3B). Given that the 2,379 *m/z* peak is detectable even at

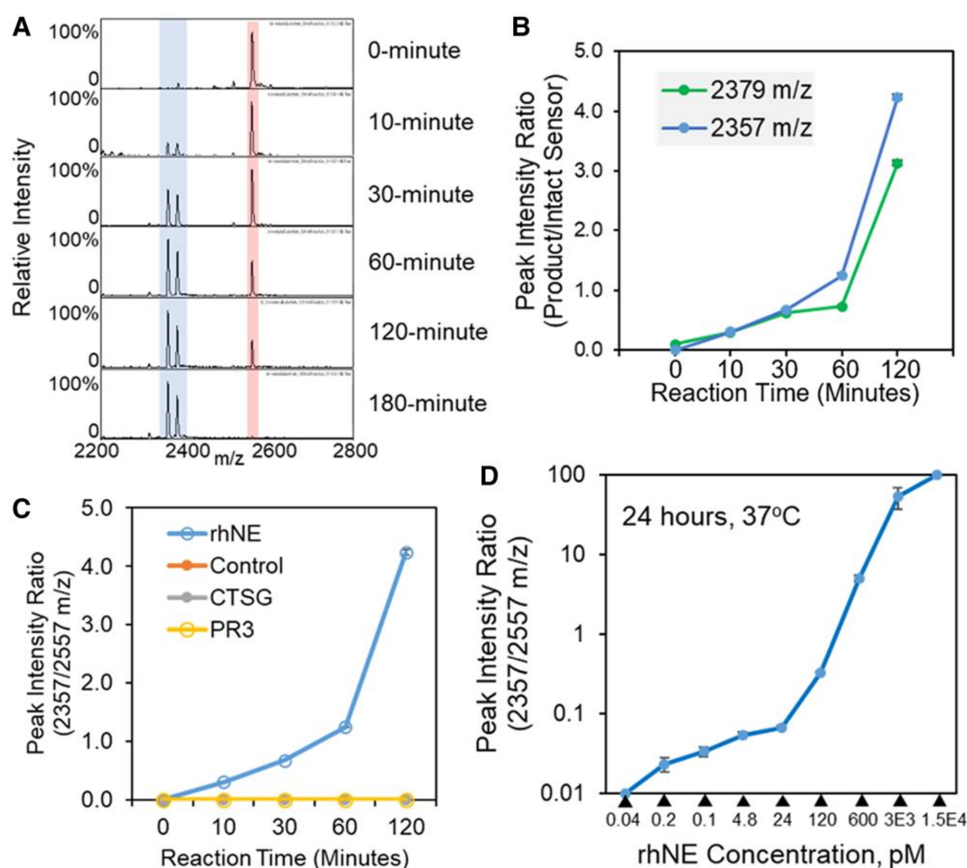

Fig. 3. Evaluation of the HNE substrate sensor using rmNE and rhNE. *A*) Representative MALDI-TOF mass spectra of the sensor and the cleavage products with different time points of reaction using rmNE. *B*) The ratio of cleavage products to the sensor with different time points. *C*) Specificity testing using a blank control, 15 μM of rhNE, 15 μM of cathepsin G (CTSG), and 15 μM of PR3. *D*) LOD test at 37 °C for 24 h using different concentrations of rhNE.

the 0-time point, potential interferences are suggested (Fig. 3A). Therefore, in our following experiments, we used the 2,357 *m/z* peak, which offers better reliability for quantitative analysis.

We then evaluated the specificity of our sensor for HNE. To this end, we incubated our sensor with a blank control sample consisting of MS-grade water and with different proteases, including 15 μM of rhNE, human CTSG, and human proteinase 3 (PR3), respectively (Fig. 3C). Our results indicated that the product mass peak at 2,357 *m/z* was only observed in the rhNE samples, suggesting that our sensor is specific to rhNE (Fig. 3C).

We constructed a quantitative calibration curve by incubating various concentrations of rhNE, ranging from 1.5E4 to 0.04 pM. We incubated the sensor with rhNE for 24 h at 37 °C (Fig. 3D). Our data indicate that the cleavage peak at 2,357 *m/z* was observed in samples with as low as 0.04 pM of rhNE (Fig. 3D). This detection capability, even at extremely low concentrations of rmNE, demonstrated the sensitivity of the customized HNE substrate sensor. Additionally, utilizing the established calibration curve from this limit of detection (LOD) analysis method, we can estimate the HNE concentration in various samples, including clinical breath samples used in this study.

The specificity of the unnatural amino acid substrate to HNE that we used in this study has been investigated in previous studies (19, 21). Chan et al. (21) designed an HNE substrate sensor using the same unnatural amino acid sequence, albeit with varied linkers and reporters, and demonstrated that rmNE and HNE showed a consistent concentration-dependent release of volatile reporters over 2 h at 37 °C. Furthermore, Chan et al. (21) conducted an evaluation of their HNE substrate sensor against various proteases, including PR3, granzyme B (GZMB), cathepsin B (CTSB), cathepsin D (CTSD), MMP9, and MMP13. Their findings revealed that the volatile reporter's release, measured by mass spectrometry, was only detected in HNE samples (21). While we did not conduct any experiments to assess this specificity, given that we follow the identical unnatural amino acid substrate for HNE, Nle(O-Bzl)-Met(O)2-Oic-Abu, we expect that our substrate sensor is HNE-specific, in alignment with previous findings.

In this study, we demonstrate that our customized HNE sensor substrate shows great sensitivity, critical for detecting ultralow protease levels in breath samples, known to have a ppb level of protein materials. Crucially, we have established a calibration curve, enabling the quantification of HNE concentrations in breath. This development demonstrates the quantitative capacity of our methodology, particularly considering previous reports revealing higher HNE activity in respiratory specimens from individuals with LRTI.

Application of the breath-based in vitro assay for the detection of LRTIs in intubated patients

Following the encouraging outcomes observed in our feasibility study, we decided to advance our research by applying this assay to human breath samples. Our breath sampling system, BreathBiomics, is currently used in the neurological ICU at JHH (Johns Hopkins Hospital, NCCU) to collect biomolecules from intubated patients' exhaled breath (Fig. 1A, Table S1) (17). Our previous studies demonstrated that this sampling system efficiently captures submicron particles originating from the respiratory tract lining fluid through the airway reopening mechanism (14, 17). Using advanced top-down proteomics, we identified rich truncated proteoforms associated with ventilator-associated pneumonia (VAP), a major LRTI in ICUs, pointing to the potential overactivation of proteases (17). Most importantly, the presence of proteases in the breath samples was identified, offering direct evidence of protease release in exhaled breath (17). In this study, we applied our in vitro assay to breath samples from intubated patients. In addition, we added a group of healthy volunteers to serve as negative controls (Fig. 1B).

The quantitative measurement of HNE activity was conducted using the calibration curve shown in Fig. 3C. For calculating HNE activity derived from this curve, cases where signals were not detected or the intensity was below the predetermined LOD were assigned a numerical value of "0". Our data revealed that the LRTI group had a significantly higher median HNE activity than the non-LRTI and healthy volunteer groups (Fig. 1C, Tables S2 and S3). The distribution range of HNE measurements in the LRTI cohort was more extensive, indicating increased variability within this group (Fig. 1C). In pairwise comparisons, the LRTI group presented a significant mean difference of 9.8 in HNE measurements when contrasted with the non-LRTI group, which was substantiated by an adjusted *P*-value of <0.001 (Fig. 1C). Similarly, a comparison between the LRTI and healthy volunteer groups revealed a significant mean difference of 9.2 in HNE concentration, confirmed by an adjusted *P*-value of <0.001 (Fig. 1C). The non-LRTI and healthy volunteer groups showed a nominal mean difference of 0.6 in their HNE measurements, which was not statistically significant as indicated by an adjusted *P*-value of 0.8 group (Fig. 1C). Representative MALDI-TOF mass spectra of each group were shown in Fig. 1D. Previous studies reported that HNE concentrations in bronchoalveolar lavage fluid (BALF) were ∼111 nM for LRTI cases. In our findings, the HNE concentration in breath samples for LRTI cases was ∼0.27 nM, suggesting that breath concentrations are significantly lower than those in BALF samples (Tables S2 and S3).

Since we demonstrated the distinct differences by our in vitro assay between LRTI and non-LRTI cases, we investigated its diagnostic potential by the quantification of HNE activity in both groups, which involved the construction of a receiver operating characteristic (ROC) and the calculation of an area under the curve (AUC). The findings of our study showed an AUC of 0.987, suggesting a robust diagnostic capacity for using HNE activity as a marker for detecting LRTI in breath samples (Fig. 1E). The best cutoff value, as defined by Youden's index for this ROC curve, is set at an HNE concentration of 0.2 pM, achieving a sensitivity of 1 and a specificity of 0.867.

Correlation between HNE activity and protein levels in breath samples

We conducted a bottom-up proteomic approach to characterize protein profiles in breath samples to investigate the potential correlation between observed HNE activity and total HNE protein content within the samples. Within these samples, the peptide VVLGAHNLSR characteristic of HNE was identified, in addition to the identification of several other proteases and antiproteases, including MMP8, MMP9, and CTSG (refer to Fig. 4A,B). Our results also showed that, among the detected proteases, HNE had the weakest intensity, suggesting a lower protein concentration relative to the other proteins identified. Subsequent correlation analysis showed a highly positive association between the quantified HNE activity and HNE protein level in the breath samples, as validated by a Spearman's rho of 0.95 and a *P*-value of <0.001 (Fig. 4C, Table S4). Such correlation was not observed for MMP8 and MMP9 (Fig. S3). Even though our assay was designed to detect protease activity, not protein level, our findings are

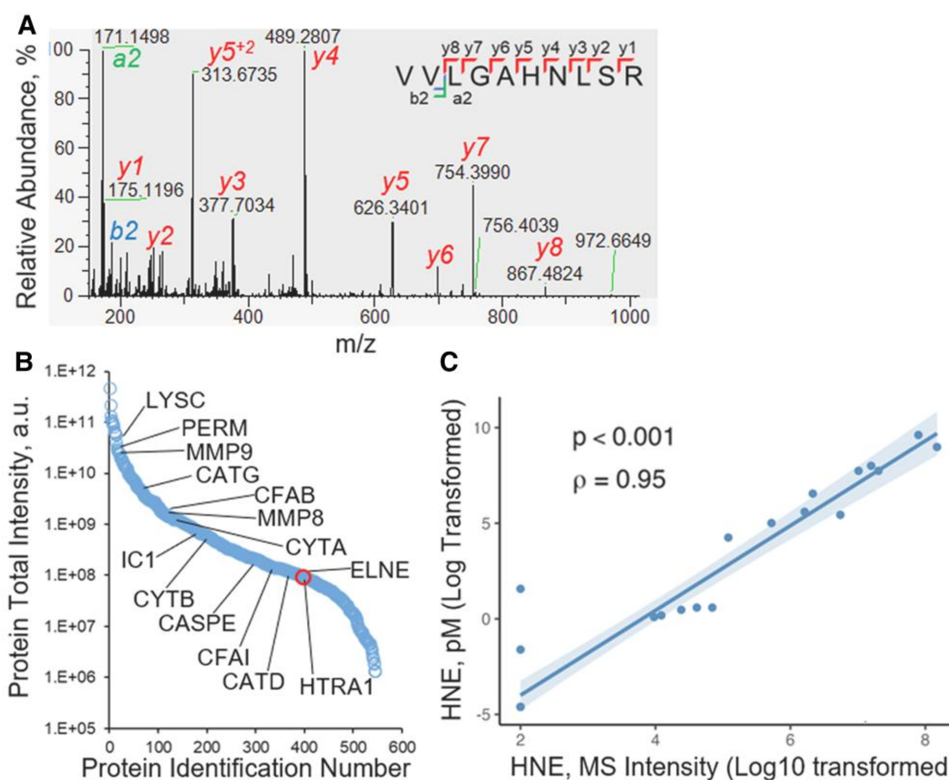

Fig. 4. Identification of HNE in breath samples using bottom-up proteomics and correction analysis. *A*) Representative ion fragmentation map of an HNE peptide. *B*) Protein profiles of proteases and antiproteases identified in breath samples of intubated patients. *C*) Correlation analysis between HNE concentrations and protein levels in breath samples.

consistent with previous studies demonstrating that the elevated HNE activity results from the overrelease of HNE protein.

The protein profiles of antiproteases and proteases in the breath samples of intubated patients were revealed by bottom-up proteomics, including lysozyme C (LYSC), myeloperoxidase (PERM), matrix metalloproteinase-9 (MMP9), CTSG, complement factor B (CFAB), neutrophil collagenase (MMP8), cystatin-A (CYTA), plasma protease C1 inhibitor (IC1), cystatin-B (CYTB), caspase-14 (CASPE), complement factor I (CFAI), CTSD, serine protease (HTRA1), and neutrophil elastase (ELNE) (Fig. 4B). The presence of multiple proteases in our breath samples implies the potential to explore additional proteases for diagnosing LRTI.

Discussion

LRTIs stand among the leading causes of mortality, particularly pronounced in intubated patients within ICUs, which is associated with a significant mortality rate, yet accurate diagnosis remains challenging, resulting in delayed treatment and overuse of antibiotics (2). Challenges in diagnosing LRTIs stem from two main factors: the absence of noninvasive methods capable of accessing the site of infection within the lower respiratory tract and the need for robust assays that can accurately measure reliable biomarkers for diagnostic purposes. In this study, we addressed these challenges by introducing an advanced sampling system designed to collect biomolecular materials in breath and establishing an in vitro assay based on the widely reported up-regulated HNE activity.

Since protease–antiprotease imbalance is a characteristic feature of infections, earlier research has proposed approaches to use this dysregulated protease activity as a diagnostic marker for the disease (8–10, 17, 21–23). Chan et al. (21) presented a breath-based assay that detected neutrophil elastase (NE) activity within animal pneumonia models by measuring the release of volatile reporters conjugated to an NE substrate nanosensor. In their experimental design, the nanosensor was administered to the lungs through intratracheal instillation. Following bacterial infection, an excess release of NE interacted with the nanosensor, leading to the release and emission of volatile reporters in the breath. These reporter molecules were then collected and analyzed using mass spectrometry (21). Contrary to our in vitro methodology, the method outlined in Chan et al.'s study involved intratracheal instillation for delivering the NE substrate nanosensor, categorizing it as an in vivo approach. Therefore, several critical limitations were discussed by the authors, including the biocompatibility and potential toxicity of both the substrate and volatile compound for human application. While the authors suggested that careful selection and design of candidate volatile reporter groups can mitigate toxicity risks, their study demonstrated a tight link between the structures of the volatile reporters and the protease-substrate interactions, showing that any structural alterations could compromise the diagnostic performance of the breath analysis (21). For example, the study demonstrated that when the volatile reporter HFA1 (hydrofluoroamine) was replaced with larger reporters, such as HFA5, there was a decrease in the catalytic efficiency of substrate cleavage by NE by as much as two orders of magnitude, which was primarily attributed to alterations at the P1′ position (21). Given that in our in vitro assay, proteases are collected from breath and subsequently incubated with the customized substrate sensor, such complications are not present in our methodology.

In Chan et al.'s (21) study, the volatile reporter signals used for monitoring respiratory diseases can be affected by the quantity of

substrate administered and absorbed within the host. In contrast, our methodology offers a unique advantage whereby both the intact substrates and their respective cleavage products are detectable and quantifiable, thereby providing a more accurate assessment of protease activity. Furthermore, we employ the MALDI-TOF mass spectrometry platform for our measurements, an instrument already prevalently incorporated within clinical environments due to its practicality and user-centric design, highlighting the potential for our method's transformation into a point-of-care device for pulmonary infection diagnostics (18). In summary, we have developed a novel in vitro methodology which detects protease activity directly in exhaled breath, successfully addressing and overcoming several critical challenges in creating a breath-based diagnostic technique for LRTIs.

Our results are consistent with the previous studies that the overrelease of HNE, one of the most abundant proteases observed in respiratory tract specimens, is a hallmark in VAP and can be used for diagnosis (8). Wilkinson et al. observed a 9.2-fold increase in HNE levels in BALF when comparing subjects with suspected VAP to control participants (8). Furthermore, HNE was typically not present in the BALF of healthy individuals. In a multicenter validation study including 12 ICUs, Hellyer et al. (10) confirmed that levels of HNE in BALF were increased by 11.1-fold in patients diagnosed with VAP compared with those without VAP. Furthermore, the use of HNE as a biomarker to discriminate between VAP cases within a cohort presenting with suspected VAP was reported in these studies, with a sensitivity range of 93–98% and a specificity range of 34–79% (10). Although previous studies demonstrated that the diagnostic performance of HNE in the critical care setting for diagnosing LRTIs was promising, the invasive nature of bronchoalveolar lavage (BAL) represents a significant limitation for its routine implementation. As demonstrated in our study, the robust performance of HNE activity as a diagnostic marker for LRTIs in human breath suggests that our noninvasive methodology can overcome the existing limitations. This advancement could significantly refine the LRTI diagnostics and improve LRTI management within critical care settings.

Recognizing the proteases associated with specific diseases is essential and necessitates the construction of a protease list for the assay (22, 23). To determine the appropriate proteases for the multiplexed assay, indirect analytical strategies focusing on understanding protease profiles in existing transactional datasets coupled with machine learning have been recommended (22). For example, Anahtar et al. (22) concluded protease signatures differentiating bacterial from viral infections using transcriptional meta-analysis incorporating public transcriptional datasets from human subjects with respective respiratory infections. Similarly, Kirkpatrick et al. (24) constructed a protease panel for human lung adenocarcinoma, comprising 15 proteases derived from an in-depth study of current human and mouse RNA-Seq and microarray datasets. Using the methodology presented in our study, we can obtain a comprehensive protease profile through noninvasive breath collection techniques. Therefore, our strategy offers a more straightforward and reliable framework for developing protease panels for multiplexed applications.

The investigation of neutrophils and associated inflammatory factors as diagnostic markers is supported by the fact that neutrophils are the frontline defense in response to pathogenic invasion and contribute to tissue damages via overreleasing extracellular traps and proteases, including MMP8, MMP9, HNE, IL-1β, IL-8, TNFα, IL-6, G-CSF, IL-10, sTREM-1, MIP-1α, and MCP-1 (7–10). Beyond the well-documented significance of neutrophil proteases in pathological inflammatory processes, two additional considerations support our selection of HNE as the target for our in vitro activity assay. First, HNE represents one of the abundant proteases secreted by neutrophils, a fact supported by neutrophil proteome analysis and, most critically, by examination of respiratory tract specimens from mechanically ventilated patients, a finding our study confirmed through the protein profiles revealed by our bottom-up proteomics approach (17, 25, 26). Specifically, the concentration of HNE in BALF samples has been quantified at ∼4,000 ng/mL (8–10). In contrast, primary cytokine levels, such as soluble TREM-1 (sTREM-1) and interleukins (ILs), are typically measured in the picogram per milliliter range, presenting a considerable sensitivity limitation for corresponding detection methods (7, 9, 10). Since protein concentrations in breath samples are substantially lower than in BALF, targeting proteases at ultralow concentrations presents a problematic challenge. Second, the performance of our assay is significantly enhanced by the previous development of an ultrasensitive substrate for HNE, synthesized from unnatural amino acids Nle(O-Bzl)-Met(O)2-Oic-Abu (19). In contrast to the traditional HNE substrate Ala-Ala-Pro-Val, this engineered substrate demonstrates a significant hydrolysis rate by HNE, ∼7,000 times faster as determined through catalytic efficiency analysis (5.8E3 vs. 4.8E7$M^{-1}.s^{-1}$, k_{cat}/K_m) (19). This significant increase in catalytic efficiency is essential for detecting subtle quantities of HNE present in our exhaled breath samples. Our study revealed similar HNE profiles with reported BALF studies, supporting the observation that HNE is typically undetectable in healthy individuals and present at low levels in the nonVAP control group (8). For example, Wilkinson et al. found that total HNE protein was virtually undetected in BALF samples from healthy subjects and significantly reduced (∼10-fold) in the VAP control groups, all with minimal sample variation (8). These are consistent with our study's findings. To summarize, from both previous BALF-based reports and our current evidence, it is clear that intubated patients with VAP exhibit significantly elevated levels of HNE protein and activity. However, the noninvasive nature of our method represents a critical advantage over invasive BAL procedures, enabling routine monitoring of intubated patients for the presence of LRTIs, a fact that distinctly underlines the value of our approach.

The pathogenesis of many diseases is determined not merely by the presence of specific proteins but, more crucially, by their functional states. Consequently, as demonstrated in our approach and others, the quantification of protease activity serves as a more accurate indicator of the pathological condition, which contrasts with methods that quantify total protein content, like immunoassays, which may not effectively discriminate between the active and inactive states of proteases. This capability to differentiate represents a distinct advantage of our approach. In our study, we observed a strong correlation between HNE total protein concentration and its protease activity, mainly attributable to HNE's release in an active form. Conversely, other proteases like matrix metalloproteinases (MMPs) are initially secreted as inactive zymogens and require extracellular enzymes for activation. Under this condition, protease activity measurement may offer more accurate diagnostics in scenarios such as bacterial infection detection. Indeed, it has been reported that MMP8 and MMP9 levels were significantly elevated in the VAP group compared with the nonVAP group (8). However, the activity of MMP9 was not significantly different between groups (8). This discrepancy between protein concentration and enzymatic activity suggests the importance of protease activity measurement in disease diagnostics (8). Moreover, pathogen infections are characterized by swift and dynamic shifts in the inflammatory response,

a phenomenon where the enzymatic activity of specific proteases may alter the progression of the disease (10). Our methodology, which focuses on the measurement of protease activity, can provide diagnostic data crucial for determining the stage and severity of the infection. This approach is particularly critical in cases where the activity of enzymes is likely to change with the progression of the infection or as a consequence of therapeutic measures. Thus, the measurement of protease activity can serve as an effective biomarker, offering insights important for the accurate diagnosis and effective management of infectious diseases.

A critical limitation within our current methodology that needs to be acknowledged is the dependence on a single protease marker, HNE, and thus lacks multiplexed analytical capacity. It has been widely accepted that multiple markers perform better than an isolated marker (21–24). This observation stems from the diversity in individual host responses to pathogenic infections, which can vary significantly from one subject to another, and the critical need to discriminate between viral and bacterial infections accurately (22). Therefore, a comprehensive diagnostic approach necessitates integrating a multiplexed strategy to contain a broader spectrum of biomarkers, thereby enhancing the performance of diagnostics. To address the clinical barriers in differentiating between viral and bacterial pneumonia etiologies, Anahtar et al. (22) presented a proof-of-concept study in which the group constructed a protease panel that included 39 proteases using data from 33 distinct and publicly accessible study cohorts to demonstrate their differential expression in bacterial versus viral respiratory infections. Subsequently, the animal study used machine learning algorithms to improve diagnostic classifiers to identify mice with pathogen infection and accurately categorize bacterial versus viral pneumonia cases. This innovative diagnostic method, thus, presents a feasible solution for determining the etiology of pneumonia, relying on the host's proteolytic response to the infection (22).

In our study, we identified a range of proteases in breath samples, including MMP8, MMP9, CTSD, and CTSG, all of which have been associated with pathogen invasions in previous reports (7–10, 21, 22). This finding suggests that our approach could be developed into a multiplexed detection system capable of designing sensors for different protease substrates. These customized sensors would be able to detect various protease activities simultaneously. Anticipating this possibility, we incorporated a MALDI-TOF MS platform, which aligns with our objectives. MALDI-TOF MS, recognized for its ease of use and rapid result delivery, is widely utilized in microbiological laboratories within hospital settings. The MALDI-TOF MS platform presents two unparalleled advantages due to its high-resolution and specificity. First, it allows accurate mass measurement of the intact substrate sensor and its cleaved product, which is crucial in minimizing errors such as missed or incorrect cleavages, thereby significantly reducing the likelihood of false positive results. Second, the high-resolution of MALDI-TOF MS facilitates the differentiation of various cleavage products and intact sensors, even with minimal mass differences. This capability enables the design and incorporation of a substantial number of protease-substrate sensors within an optimal mass range. Based on our understanding and estimations, the platform could potentially adapt the detection of as many as 80 different protease substrates. Incorporating the MALDI-TOF MS platform in our approach represents a strategic advancement in improving our diagnostic capabilities, especially in the multiplexed capacity that allows for the simultaneous assessment of multiple proteases with accuracy, enabling robust diagnostics. It should be noted that traditional diagnostic methods often focus on either the detection of pathogen markers or human response markers. However, an integrated approach that includes both is crucial for enhancing diagnostic accuracy (27, 28). Additionally, a well-controlled study that confirms antibiotic use and evaluates its impact on the results of our platform would be valuable, which would provide insights into how antibiotic treatment influences the diagnostic accuracy and effectiveness of the BreathBiomics platform, providing more comprehensive data for clinical decision-making.

In summary, we introduced a breath-based methodology for detecting LRTIs, a significant step forward in using noninvasive analysis for disease diagnosis. We wanted to acknowledge that this is a proof-of-concept study, with substantial further work to establish diagnostic performance in a population with clinical suspicion of infection—where it would be deployed. Our study provides strong evidence supporting the correlation between HNE activity and LRTIs, demonstrating significant diagnostic promise. We anticipate that our approach could serve as a valuable screening tool for LRTIs, particularly in intubated patients within critical care settings. Following rigorous validation in clinical studies, this technology holds substantial promise for aiding physicians in clinical decision-making processes, including the diagnosis, guiding antibiotic treatment strategies, and overall management of LRTI conditions, thereby marking a significant advancement in medical diagnostics.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

This work was supported by the National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH) under Awards R44AI177245.

Author Contributions

D.C.: Conceptualization, Funding acquisition, Methodology, Supervision, Writing original draft, Writing reviwer & Editing. M.A.M.: Methodology. S.C.: Methodology. W.A.B.: Methodology. M.M.: Methodology. K.M.K.: Investigation. E.R.C.: Investigation. C.R.H.: Investigation. M.S.C.: Investigation. Y.P.: Data curation.

Data Availability

The data in this study are included in the article and the supplemental file.

References

1 Reddy KS. 2016. Global Burden of Disease Study 2015 provides GPS for global health 2030. *Lancet*. 388(10053):1448–1449.
2 Kalil AC, *et al.* 2016. Management of adults with hospital-acquired and ventilator-associated pneumonia: 2016 clinical practice guidelines by the Infectious Diseases Society of America and the American Thoracic Society. *Clin Infect Dis*. 63(5):e61–e111.
3 Papan C, *et al.* 2018. Assessment of the multiplex PCR-based assay Unyvero pneumonia application for detection of bacterial pathogens and antibiotic resistance genes in children and neonates. *Infection*. 46:189–196.

4 Locher K, *et al.* 2019. FilmArray respiratory panel assay: an effective method for detecting viral and atypical bacterial pathogens in bronchoscopy specimens. *Diagn Microbiol Infect Dis*. 95(4):114880.
5 Bonten MJ, Gaillard CA, de Leeuw PW, Stobberingh EE. 1997. Role of colonization of the upper intestinal tract in the pathogenesis of ventilator-associated pneumonia. *Clin Infect Dis*. 24(3):309–319.
6 Gibot S, *et al.* 2004. Soluble triggering receptor expressed on myeloid cells and the diagnosis of pneumonia. *N Engl J Med*. 350(5): 451–458.
7 Morris AC, *et al.* 2010. Diagnostic importance of pulmonary interleukin-1β and interleukin-8 in ventilator-associated pneumonia. *Thorax*. 65(3):201–207.
8 Wilkinson TS, *et al.* 2012. Ventilator-associated pneumonia is characterized by excessive release of neutrophil proteases in the lung. *Chest*. 142(6):1425–1432.
9 Hellyer TP, *et al.* 2015. Diagnostic accuracy of pulmonary host inflammatory mediators in the exclusion of ventilator-acquired pneumonia. *Thorax*. 70(1):41–47.
10 Hellyer TP, *et al.* 2020. Biomarker-guided antibiotic stewardship in suspected ventilator-associated pneumonia (VAPrapid2): a randomised controlled trial and process evaluation. *Lancet Respir Med*. 8(2):182–191.
11 Coelho L, *et al.* 2018. C-reactive protein and procalcitonin profile in ventilator-associated lower respiratory infections. *J Crit Care*. 48:385–389.
12 Bake B, Larsson P, Ljungkvist G, Ljungström E, Olin AC. 2019. Exhaled particles and small airways. *Respir Res*. 20:1–4.
13 Behndig AF, Mirgorodskaya E, Blomberg A, Olin AC. 2019. Surfactant protein A in particles in exhaled air (PExA), bronchial lavage and bronchial wash - a methodological comparison. *Respir Res*. 20(1):1–6.
14 Chen D, Bryden WA, McLoughlin M. 2020. A novel system for the comprehensive collection of nonvolatile molecules from human exhaled breath. *J Breath Res*. 15(1):016001.
15 Viklund E, *et al.* 2022. Severe acute respiratory syndrome coronavirus 2 can be detected in exhaled aerosol sampled during a few minutes of breathing or coughing. *Influenza Other Respir Viruses*. 16(3):402–410.
16 Holz O, Müller M, Carstensen S, Olin AC, Hohlfeld JM. 2022. Inflammatory cytokines can be monitored in exhaled breath particles following segmental and inhalation endotoxin challenge in healthy volunteers. *Sci Rep*. 12(1):5620.
17 Chen D, *et al.* 2023. Human exhaled air diagnostic markers for respiratory tract infections in subjects receiving mechanical ventilation. *J Breath Res*. 17(2):026001.
18 Welker M, Van Belkum A, Girard V, Charrier JP, Pincus D. 2019. An update on the routine application of MALDI-TOF MS in clinical microbiology. *Expert Rev Proteomics*. 16(8):695–710.
19 Kasperkiewicz P, *et al.* 2014. Design of ultrasensitive probes for human neutrophil elastase through hybrid combinatorial substrate library profiling. *Proc Natl Acad Sci U S A*. 111(7):2518–2523.
20 Chen D, *et al.* 2018. Top-down analysis of branched proteins using mass spectrometry. *Anal Chem*. 90(6):4032–4038.
21 Chan LW, *et al.* 2020. Engineering synthetic breath biomarkers for respiratory disease. *Nat Nanotechnol*. 15(9):792–800.
22 Anahtar M, *et al.* 2022. Host protease activity classifies pneumonia etiology. *Proc Natl Acad Sci U S A*. 119(25):e2121778119.
23 Soleimany AP, Bhatia SN. 2020. Activity-based diagnostics: an emerging paradigm for disease detection and monitoring. *Trends Mol Med*. 26(5):450–468.
24 Kirkpatrick JD, *et al.* 2020. Urinary detection of lung cancer in mice via noninvasive pulmonary protease profiling. *Sci Transl Med*. 12(537):eaaw0262.
25 Nguyen EV, *et al.* 2013. Proteomic profiling of bronchoalveolar lavage fluid in critically ill patients with ventilator-associated pneumonia. *PLoS One*. 8(3):e58782.
26 Kotz KT, *et al.* 2010. Clinical microfluidics for neutrophil genomics and proteomics. *Nat Med*. 16(9):1042–1047.
27 Alsved M, *et al.* 2023. Size distribution of exhaled aerosol particles containing SARS-CoV-2 RNA. *Infect Dis*. 55(2):158–163.
28 Kurver L, *et al.* 2022. SARS-CoV-2 RNA in exhaled air of hospitalized COVID-19 patients. *Sci Rep*. 12(1):8991.