Protease activated nanosensors for the noninvasive diagnosis of community-acquired pneumonia

by

Melodi N. Anahtar

B.S., Massachusetts Institute of Technology (2016)

Submitted to the Harvard-MIT Division in Health Sciences and Technology
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Medical Engineering and Medical Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2022

Author . . . . . .
Harvard-MIT Division in Health Sciences and Technology
December 17, 2021

Certified by . . . . . .
Sangeeta Bhatia, MD, PhD
Wilson Professor of Health Sciences and Technology
Electrical Engineering and Computer Science, MIT
Thesis Supervisor

Accepted by . . . . . .
Emery N. Brown, MD, PhD
Director, Harvard-MIT Program in Health Sciences and Technology
Professor of Computational Neuroscience and Health Sciences and Technology

Protease activated nanosensors for the noninvasive diagnosis of community-acquired pneumonia

by
Melodi Anahtar

Submitted to the Harvard-MIT Program in Health Sciences and Technology
on December 17, 2021 in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy

Abstract

Community-acquired pneumonia is the most common infectious cause of death worldwide, and is defined as an infection of the lung parenchyma. It is a heterogeneous disease caused by wide range of bacteria, viruses, and occasionally fungi. An arsenal of diagnostic tools have been developed to detect the many causes of CAP, yet in approximately 60% of CAP patients the causative organism is never determined. As such, the standard of practice for treating suspected CAP is to administer antibiotics as soon as possible because existing diagnostics cannot determine the etiology of disease quickly and accurately enough to warrant withholding treatment. This diagnostic and therapeutic paradigm is out of touch with the approaches to personalized medicine that are being taken for other diseases such as cancer, and an urgent need exists for an accurate and rapid pneumonia diagnostic that can simultaneously detect CAP and stratify etiology.

To address this gap, in this work we have developed two novel approaches to diagnosing CAP. Both approaches leverage differential protease expression by the host in response to pneumonia-causing pathogens to detect pneumonia and stratify etiology. To this end, we first derived a 40-gene signature from human transcriptomic data, which consisted of proteases biomarkers for pneumonia. We then used our lab's activity-based nanosensor (ABN) technology to create a 20-plex panel of nanoparticles that could produce urinary signatures of disease state in response to the activity of a subset of these proteases. We validated that this panel could generate unique urinary signatures of disease in five *in vivo* mouse models of CAP within two hours of sensor administration. Using these signatures, we trained diagnostic classifiers to distinguish healthy mice from those with bacterial and viral pneumonia with high accuracy. To produce an even faster readout, we then modified these ABNs with volatile reporters to create breath-based volatile activity-based nanosensors (vABNs), and demonstrated that we that could detect pneumonia within 15 minutes of administration. Altogether, these nanosensors enable urine and breath-based detection of CAP, and constitute a means of diagnosing pneumonia that is orthogonal to existing clinical tests, thus opening a new direction of study for pneumonia diagnostics.

Thesis Supervisor: Sangeeta N. Bhatia, M.D., Ph.D.
Title: John J. and Dorothy Wilson Professor of Health Sciences and Technology &
Electrical Engineering and Computer Science, MIT

Acknowledgments

First, I'd like to thank all of the mice that gave their lives for this work. Especially Lucinda, who was my favorite mouse of all. This is not sarcastic. Second, to my committee, thank you for all of your suggestions, support, and guidance, which helped shaped this work into a thesis-worthy form. And third, thank you to Dr. Sahil Patel, Dr. Tahoura Samad and Dr. Lian-Ee Ch'ng for reading over parts of this dissertation and providing thoughtful edits and feedback.

To the members of LMRT, thank you all for the endless support over the past five years. I'd especially like to thank a few people. First, Sangeeta, whose guidance and understanding made all of this work possible. I can think of very few people, let alone PIs, who would have encouraged their trainee to step away from the bench for months to get their head on straight. I am so grateful that you never lost faith in me, even when I wasn't sure in myself. To Heather, who was always willing to lend an ear whenever I was having a problem, no matter how big or small. To Lian-Ee, my bench buddy, fellow dog lover, and the keeper of the lab without whom none of the experiments in this thesis would have been executed. To Leslie, who showed me the ropes about how to design experiments, work with mice, manage projects, and so much more. I'm so grateful for your mentorship. And of course, to all the members of the lab past and present, for their mentorship, friendship, and support. Special shout outs to Jesse, Carmen, Tahoura, Henry, Kelsey, Keval, Colin, Quinton, Liang, and Chay, for everything.

To my friends, who made life over the past few years all kinds of fun. My lab pals to life pals – Niketa, my work-wife/daughter/sister, and Ava, my mentor/walk buddy/constant outlet for all of my random thoughts. My forever roommates Julia and Catherine, who made sure that I had a life outside of lab that kept me fulfilled and loved even when everything was going wrong. My cohort, who were the reason I stayed at MIT for grad school and made it a fun ride; especially The Ladies and our Jay, who made Thursdays my favorite day of the week.

To my family, especially my mom, dad, and sister, I love you guys so much!!! I'm especially thankful that my sister created my favorite tiny human, Vera, who was the world's cutest distraction from lab and the funniest giver of life advice. And finally to my chosen family, my boyz, my quaranteam, my two best buds in the whole entire universe, Nathan and Duke, who were the least expected and best part of the last five years.

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

List of Figures

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

1.1 Causes and detection of community-acquired pneumonia

Hippocrates first described pneumonia (Greek: $\pi\nu\epsilon\upsilon\mu o\nu\iota\alpha$) in the fourth to fifth century BC. [1] This was over 2,000 years before Antoni van Leeuwenhoek's discovery of bacteria (then termed "animalcules", in 1676) and Dmitiri Ivanovsky's identification of viruses (Tobacco mosaic virus, in 1892). Without knowing what agents caused it, Hippocrates correctly identified that patients with pneumonia had liquid-filled lungs accompanied by fever, chills, coughing, pain, and difficulty breathing. [1] Since those early days, pneumonia has continued to plague humanity, killing millions of adults and children worldwide each year.

1.1.1 Defining community-acquired pneumonia

Each day, we breathe in millions of aerosolized pathogens. [2] The robust immune system in the lungs is constantly at work preventing these foreign microbes from causing respiratory infections. These infections can occur in both the upper and lower respiratory tracts, which each play unique yet interdependent roles in preventing disease. To broadly summarize their interaction, the upper respiratory tract (URT, consisting of the nasal cavity, larynx and

pharynx) keeps foreign pathogens and particles from entering the lower respiratory tract (LRT, made up of the trachea, bronchial tree, and lungs), where infections and resulting inflammation can acutely impact gas exchange and impair respiratory mechanics, thereby compromising breathing. As a result, lower respiratory tract infections (LRTIs) have particularly dire health consequences and were the fourth leading cause of death worldwide in 2019. [3]

While the term "LRTI" is sometimes used as a synonym for pneumonia, LRTI is a broader term that encompasses not only infectious pneumonia, but also bronchitis and pulmonary abscess. The distinction is that pneumonia is an infection of the lung parenchyma, but the nomenclature surrounding pneumonia and other LRTIs is subtle. Furthermore, infectious pneumonia can be further classified clinically into four broad groupings: community-acquired, hospital-acquired pneumonia, ventilator-associated pneumonia (VAP) and aspiration pneumonia. [4] As the names suggest, hospital-acquired pneumonia (HAP) is characterized as pneumonia that develops in a healthcare setting and presents at least 48 hours after hospitalization, whereas community-acquired pneumonia (CAP) develops outside of a hospital or within 48 hours after admission. CAP is generally more relevant to public health as it is dependent on community spread, and as such it will be the focus of this work.

1.1.2 Common etiologies of CAP

The prevention, diagnosis, and treatment of CAP is complicated by the fact that a wide range of bacteria, viruses, and fungi can cause disease. Not only that, but the pneumonia-causing microbe cannot be identified in a majority of patients, making it impossible to fully understand the scope of etiologies. [5] This is a relatively recent dilemma as before 1945, a single bacterium, *Streptococcus pneumoniae*, was responsible for over 90% of pneumonia cases in adults. [6] Since then, the prevalence of *S. pneumoniae* in the United States has been steadily declining, a trend that has been largely attributed to the use of antibiotics and widespread vaccination of the pneumococcal polysaccharide vaccine. [6,7] This decrease in pneumococcal pneumonia has been accompanied by increased detection of other bacterial

etiologies, namely *Mycoplasma pneumoniae*, *Haemophilus influenzae*, and *Staphylococcus aureus*. [7,8] Importantly, the bacteria responsible for CAP differ based on elements such as geography, age, time of year, and individual patient risk factors. [8] For example, gram negative bacteria such as *Klebsiella pneumoniae* and *Burkholderia pseudomallei* are common causes of CAP throughout Asia but are much rarer in the United States. [9] Patient outcomes for those with bacterial CAP are improved with early initiation of antibiotic therapy, which has led to a sense of urgency among clinicians to begin treatment with empiric antibiotics as soon as CAP is suspected. [10] However, this has led to overtreatment in patients whose CAP is actually viral in origin, which can contribute to antibiotic resistance. [11]

Despite historical emphasis on bacterial CAP, there is increasing evidence that viruses are a significant cause of CAP. [6,7,12–14] In the United States, rhinovirus, influenza, and metapneumovirus are the leading viral etiologies [8], but the full scope of viral CAP is impossible to appreciate, as detecting viruses via assays such as polymerase chain reaction (PCR) has only recently been made accessible enough to enable routine testing of patients in high and middle income countries. Furthermore the exquisite sensitivity of PCR also makes it prone to false positives, because pathogen-derived nucleic acids are not necessarily derived from pathogens that are responsible for acute illness (e.g. commensal bacteria and lingering viral RNA after infection resolution). [15,16] Viral antigen tests and viral cultures can also be used to identify viral CAP, but suffer from limitations that are causing them to be edged out by PCR. [17] Namely, viral antigen tests only exist for a subset of viruses (e.g. influenza and SARS-CoV-2), limiting their utility in the broader scope of CAP diagnosis.

1.2 The diagnostic paradigm for CAP

To diagnose CAP, clinical guidelines recommend that patients with signs of a respiratory infection (e.g., cough, dyspnea, fever) be given chest X-rays to look for signs of fluid build up and inflammation in the lungs (e.g., consolidation, infiltrates, or cavitations). [18] Positive chest imaging and the presence of symptoms are typically enough to establish CAP as

a differential diagnosis, but indications of infection using routine blood tests (e.g., presence of leukocytosis or increased neutrophilic bands indicating "left shift") can be used to support the diagnosis. Once a patient is diagnosed with CAP, their severity is then assessed. There are formal scoring systems such as the CURB-65 score and the Pneumonia Severity Index (PSI) that use patient characteristics such as age, the presence of co-morbidities, and lab findings to distinguish mild from severe disease; but the decision of whether to admit a patient is usually made on a case-by-case basis based on signs, symptoms, and test results. Regardless of whether a patient is treated as an outpatient or is hospitalized, standard practice is to initiate empiric antibiotics in any patient with suspected CAP and monitor their response to therapy for improvement within 48-72 hours. [19]

Notably, this diagnostic paradigm does not include microbiologic testing for the cause of one's CAP. Depending on the season and regional spread, patients may be tested for influenza via a rapid antigen test or PCR, and since the Covid-19 pandemic, testing for SARS-CoV-2 has also become routine. Otherwise, clinical guidelines advise against microbiologic testing in otherwise healthy outpatients, including sputum cultures, because it could delay antibiotic therapy. [19] This means that patients whose CAP is due to viral causes besides flu or SARS-CoV-2 can readily be prescribed antibiotics. While these antibiotics could be effective in preventing secondary bacterial infections, which are common with primary viral CAP, as an initial treatment these antibiotics would prove to be ineffective in treating a patient's underlying illness and could give rise to antibiotic resistance. Not only that, but the one recommended diagnostic test, chest imaging, is both unable to stratify etiology [20] and often unavailable in primary care clinics, meaning that patients who are not treated at a hospital, specialized imaging facility, or a similarly equipped clinical setting are often diagnosed and treated without the radiographic proof that is the cornerstone of this diagnostic workflow. In defense of the clinical guidelines, even in hospitalized patients who are given extensive microbiological testing, including blood and sputum cultures, urine antigen testing, and multiplexed molecular assays, a causative pathogen is unable to be identified in approximately 60% of cases. [8]

Altogether, these issues highlight the fact that in an era of personalized medicine and advanced biotechnology, the current methods of diagnosing CAP are insufficient. Towards this end, new biomarkers and technologies are being studied to more accurately diagnose pneumonia and reduce the time of pathogen identification, potentially to the point where microbiologic testing can be used to guide antibiotic stewardship.

1.2.1 Serum biomarkers for distinguishing CAP etiology

In hospitals, serum biomarkers such as C-reactive protein (CRP) and procalcitonin (PCT) are used to broadly distinguish between bacterial and viral infections. CRP is synthesized by hepatocytes in response to inflammation and tissue damage, and high levels have been associated with bacterial infections. [21, 22] The Alere Afinion CRP, Eurolyser Smart, and QuikRead Go CRP are all devices that can be used to rapidly quantify CRP levels in the blood. However, even if a patient has CRP levels indicative of bacterial pneumonia, these tests do not specify the disease-causing organism, greatly limiting their utility. PCT is a polypeptide that becomes markedly elevated in patients with bacterial infections and septicemia. [23] There is growing evidence that it is a better indicator of bacterial vs. viral infection compared to CRP. [24] However, both CRP and PCT ultimately suffer from poor sensitivity and have been shown to be insufficient for diagnosing CAP, let alone distinguish between bacterial and viral causes. [25] Furthermore, the thresholds for differentiating bacterial versus viral pneumonia using these markers are still being debated, and are ultimately not reliable enough to broadly influence decision-making on antibiotics. [26] Hence, there is still a need for specific, reliable CAP biomarkers.

To this end, much research is being done to find novel biomarkers that can distinguish etiology. Often, these candidates are first identified as sepsis biomarkers, and are subsequently evaluated for their ability to distinguish CAP etiology. Several examples include mid regional pro-adrenomedullin (MR-proADM), which is a more stable form of adrenomedullin (which functions as a vasodilator, diuretic, and bronchodilator), the ratio of Fas apoptotic inhibitory molecule 3 to placenta-specific 8 (FAIM3:PLAC8), and calprotectin, a protein

derived from the cytosol of neutrophil granulocytes [27–29]. While biomarkers such as these represent a step forward in CAP diagnosis, they all rely on detection from blood. Therefore, they will suffer from the classic limitations of all blood biomarkers, such as the need to establish threshold values for a positive test without sacrificing specificity, inherent dilution in the 5-6 liters of circulating blood, and the lack of site-specific signal as blood contains molecules that are derived from all over the body.

1.2.2 Multiplexed nucleic acid amplification tests for identifying CAP-causing organisms

Issues with sensitivity and specificity can be overcome using multiplexed, automated PCR panels, which have been developed to detect bacterial and viral nucleic acids from a wide range of samples. For example, the BioFire FilmArray Pneumonia (PN) Plus Panel can detect 9 viruses, 18 bacteria (including common lung colonizers), and 7 antimicrobial resistance genes from bronchoalveolar lavage (BAL), endotracheal aspirate, and sputum samples in approximately 1 hour, which is far superior to traditional culture and PCR methods. BioFire also produces other FilmArray respiratory and blood culture panels that can expand the platform's detection repertoire. A wide array of other panels such as the Curetis Unyvero LRT panel and the SeeGene Allplex™ Respiratory Panel Assays also enable multiplexed PCR. In general, these panels have excellent performance in clinical settings [30, 31], and have been shown to decrease the time course of antibiotic therapies. [32–34]. While these panels are fast and accurate, they require users to use proprietary instrumentation or software, making these tests untenable for use in many outpatient and resource limited settings. For example, the BioFire panels can only be run using the BioFire FilmArray instruments, which cost tens of thousands of dollars. In addition, each test needs to be performed on a disposable cartridge, which are around $120 each, making the cost to use and maintain such a system prohibitive for many outpatient facilities where patients with mild CAP are likely to present. [35]

Furthermore, the samples needed to run these panels and similar tests are difficult to

collect. For one, many CAP patients do not naturally produce sputum, necessitating induction that can only be performed in patients with high enough oxygen concentrations who can protect their airway, and even then getting quality samples is notoriously difficult. [36] Studies have shown that sputum can be reliably produced through the nebulization of hypertonic saline [37], but induced sputum can easily be contaminated by the microflora of the URT and have a composition that is inadequate for testing. Similarly, bronchoscopy with BAL is an invasive procedure that requires sedation in order to instill saline into the lungs and suction it back out with a bronchoscope. While BAL can overcome some of the limitations of sputum by allowing for high quality lower respiratory tract microbial samples in patients regardless of their ability to produce sputum, it has several of its own limitations including the need for an invasive procedure with risk for procedural complications, need for trained personnel, and expense. [37]

1.2.3 Urine antigen tests for identifying CAP-causing organisms

A far less invasive method of diagnosing CAP is through urine antigen tests for *S. pneumoniae* and *Legionella*. The Alere BinaxNOW *S. pneumoniae* test detects C-polysaccharide, which is present in all strains but not all serotypes, within 15 minutes. Similarly Alere BinaxNOW *Legionella* Urinary Antigen Card can detect serogroup 1 bacteria, which is responsible for a majority of human infections, but is not all encompassing. While these tests are fast and clinically useful for select organisms, nationwide CAP treatment guidelines discourage their routine use unless there is high epidemiological risk of these bacteria or the patient is suffering from severe CAP, as there is conflicting evidence about whether their use influences antibiotic treatment and patient outcomes. [19, 38, 39] Furthermore, these urine tests have been shown to have few benefits in terms of cost saving. [40]

1.3 The role of proteases in pneumonia

Much like cancer, pneumonia is not a single disease, but a disease state that is characterized by a common underlying condition: an acute infection of the lung. It was once believed that the presence of microorganisms in the LRT sullied an otherwise sterile lung environment, sparking pneumonia. [41] However, it is now known that the healthy LRT contains a wide range of colonizing bacteria that help maintain homeostasis, meaning it not simply the presence of microorganisms that triggers infection. Rather, this homeostasis can be disrupted by the host response to foreign microbes and by the pathogens themselves through the production of virulence factors. [8] The lungs prevent such disruptions from leading to pneumonia throughout the balance of two forces: immune resistance and tissue resilience. [42] Immune resistance involves killing the pathogenic microbes, whereas tissue resilience encompasses the processes of maintaining or restoring the structural and functional integrity of the lung despite the surging immune response. Those who study these processes typically focus on the cytokines, chemokines, receptors, and signaling molecules that are involved. In this work, we are approaching respiratory immunity through the lens of protease activity as a driver for immune resistance and tissue resilience in response to pneumonia.

Proteases are enzymes that degrade proteins. There are over 500 proteases produced by humans, and an even greater number produced by other organisms, including bacteria and viruses. [43] There are five classes of proteases (serine, cysteine, metallo, aspartic and threonine) that are distinguished by the catalytic mechanisms they use to hydrolyze peptide bonds. Serine, cysteine, and metalloproteases are the most abundant, each making up approximately 30% of identified human proteases, respectively. [44] Proteases have been well studied in the contexts of the coagulation cascade, where they activate platelets and other proteases, and cancer progression, where they play significant roles in tissue remodeling and metastasis. [45] This work will focus on proteases that are produced by the host to combat infection through the processes of immune resistance and tissue resilience.

When considering pneumonia that is caused by a pathogen that the body has not encountered before, the processes of innate pulmonary immunity become particularly important.

Traditionally, research into the innate immune system of the lower respiratory tract emphasizes the important of the lung epithelium, which is the surface directly in contact with inhaled air, and innate immune cells, which are the first responders to infection. [46] The alveolar epithelium is lined by Type I and Type II lung cells, which jointly serve as the lung's protective barrier against pathogens and maintain the gas exchange barrier. Type II pneumocytes, specifically, maintain lung structure by producing surfactant, containing the proteins SP-A and SP-D, which are crucial in supporting the innate immune response in the lungs and help prevent infection by binding to microbial surfaces via collectins and lipopolysaccharide (LPS), thereby priming them for clearance by alveolar macrophages (AMs). In the healthy lung environment, AMs dominate the immune cell population in the respiratory tract, where they play an immunosuppressive role by clearing debris, excess surfactant, and apoptotic cells. [47, 48] However, in the setting of infection AMs are pro-inflammatory, phagocytosing macrophages and triggering cytokines that help orchestrate the immune response. [42, 48] This phagocytosis is supported by the production of cathepsins, which are lysosomal proteases that aid in killing phagocytosed bacteria via apoptosis induction. [49] AMs also produce matrix metalloproteases (MMPs), which process chemokines, remodel the extracellular matrix, and generally play important roles in inflammation. [49]

Another important immune cell involved in pulmonary immunity is the neutrophil. Unlike AMs, neutrophils have a very small population in healthy lungs, but after infection, they are the most rapidly and abundantly recruited cells to the disease site. [42] This recruitment is orchestrated by chemokines, which cause neutrophils to migrate to the infected lung and immediately take on several effector (e.g., degranulation, production of reactive oxygen species and neutrophil extracellular traps, phagocytosis), and inflammatory (e.g., chemokine and cytokine signaling to recruit other neutrophils and immune cells to the site of infection) roles. Particularly relevant to this work are the neutrophil serine proteases (NSPs) that are released via neutrophil degranulation, including neutrophil elastase (NE), proteinase 3 (PR3) and cathepsin G (CTSG). It has been shown that these proteases play a direct role in killing intracellular *S. pneumoniae*. [50] However, harkening back to the

idea of tissue resilience versus immune resistance in maintaining lung homeostasis, this proteolytic action is a double-edged sword as *S. pneumoniae* also produces pneumolysin, a virulence factor that induces the neutrophil lysis, thereby releasing these proteases into the extracellular environment. This results in proteolytic damage to the alveolar epithelium and macrophages. [51] *S. aureus* has also developed its own protective mechanism against NSPs through the production of extracellular adherence proteins, which inhibit NSP activity thereby reducing bacterial killing and enabling further colonization.

Protease release via degranulation is also a hallmark process for natural killer (NK) and cytotoxic T lymphocytes (CTLs), which produce granzymes in response to infection. While NSPs are associated with bacterial infections, granzymes are typically associated with viral infections, as NK and CTLs are highly implicated in viral immunity. [52] In contrast to the previously mentioned proteases that directly attack the bacteria, granzymes trigger apoptosis in infected cells thereby preventing further spread. [**?**] This mechanism of action accounts for the important role of granzymes in viral infections, as viruses hijack host cells and require their machinery in order to survive. Overall, it is clear that proteases play an important role in pneumonia pathogenesis. In this work, we argue that by querying protease activity, it is possible to detect pneumonia-associated host immune responses.

1.4 Thesis overview

Over the past several years, the Bhatia lab has developed nanoparticles that respond to protease activity. These activity-based nanosensors (ABNs) consist of a core (e.g. iron oxide or polyethylene glycol) that is conjugated to peptide substrates. These peptides contain a reporter, such that when the substrate comes in contact with its target protease, the reporter is released from the ABN. Typically, this reporter is a small molecule that can be filtered by the kidney and concentrated in the urine. Upon urine collection, the reporter can be detected by a number of modalities. For example, if the ABNs are created with fluorescent reporters, detecting protease activity is as simple as measuring the fluorescence of collected urine. [53]

In some cases the reporters are designed to be mass-encoded, meaning they have slightly different molecular weights, which can be differentiated using mass spectrometry. [54–56] A major asset of ABNs is that they allow for direct monitoring of the disease state within the host and can be used to monitor disease progress, assuming that protease levels vary based on the patient's condition. This system is also capable of multiplexing by co-administering ABNs that are responsive to different proteases and release unique reporters, allowing a reporter signature to be generated that is indicative of a particular disease state.

Given the gaps in the CAP clinical workflow, the goal of this thesis is to create ABNs that can detect CAP and stratify etiology quickly and non-invasively. Towards this end, in **chapter 2** we designed a 20-plex panel for pneumonia based on human transcriptomic data, validated and assessed its targets *in vitro*, and showed that it produced unique urinary signatures *in vivo*. In **chapter 3**, we then showed that these urine outputs could be used to train diagnostic classifiers that could detect pneumonia and stratify etiology with high accuracy, and tested the limits of these classifiers on minimal training sets. We then created a novel readout of our ABN system by modifying the nanosensors to generate a synthetic volatile signal, enabling us to detect pneumonia via breath. In **chapter 4**, we describe the creation of these volatile activity-based nanosensors (vABNs), the development of an *in silico* model that can be used to enhance the design and predict the output of vABNs, and validated their efficacy in mouse models of pneumonia. Overall, this work demonstrates a novel approach to diagnosing pneumonia and identifying etiology that is distinct from all existing methods, is relatively fast, and uniquely non-invasive, thus opening up a new avenue of pneumonia diagnostics.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 2

Creating a urine-based readout of pneumonia by leveraging differential protease activity

2.1 Introduction

At the start of 2020, the world watched as the novel severe acute respiratory syndrome coronavirus (SARS-CoV-2) began spreading around the world, leaving few communities untouched by coronavirus disease 2019 (COVID-19). This new form of community acquired pneumonia (CAP) put pneumonia at the forefront of both medical research and public discourse. However, CAP has caused significant morbidity and mortality for thousands of years, with millions of people affected globally and over 100,000 deaths per year in the U.S. alone. [9] The COVID-19 pandemic has simply underscored how important it is to determine the etiology of one's pneumonia for the sake of individual patient management and public health. Thankfully, the world watched in awe as new molecular diagnostics were rapidly created to detect SARS-CoV-2 with exquisite sensitivity and specificity, and were quickly deployed via emergency use authorization. Unfortunately, determining the etiology of CAP in cases where SARS-CoV-2 is not the culprit is still fraught with difficulty, and

oftentimes the causative organism of someone's CAP is never identified for reasons that are unknown. [5,57] Clinical symptoms and radiologic parameters have poor specificity for distinguishing between common bacterial and viral causes. [19] As a result, the standard of care for patients with suspected CAP is to initiate empiric antibiotics as soon as possible based on local antibiotic resistance patterns and patient characteristics (e.g. age, comorbidities). [58] To accurately triage, treat, and track patients with CAP due to bacterial and viral causes, new noninvasive tools need to be developed that can both rapidly diagnose acute pneumonia and identify the etiology.

To this end, host response-based disease signatures have been developed that leverage the differential expression of host gene sets to distinguish bacterial versus viral infections. [59] These signatures mainly consist of genes encoding inflammatory markers, transmembrane proteins, and binding proteins that are implicated in the host immune response to infection but can only be made clinically useful by measuring their relative abundance through methods such as blood tests and gene sequencing. Meanwhile, a potential trove of biomarkers that can generate a functional readout of their own activation lies within the over 550 human proteases that respond to, cause, and manage disease. [44] Proteases are widely known for their roles in cancer, vascular disease, apoptosis, and inflammation. However, they are becoming increasingly recognized as potential therapeutic targets for infectious disease due to their involvement in the immune response and their production by invading pathogens. [60, 61] These same properties make them excellent candidates as biomarkers, and are made all the more appealing by the fact that their enzymatic activity enables detection by means other than simply measuring their absolute concentrations in blood. [62–64]

Our lab has previously developed activity-based nanosensors (ABNs), which are nanoparticles that can detect protease activity *in vivo* and create urinary signatures of active disease. [53–56, 65–68] These ABNs contain mass-encoded peptide linkers that are designed to be cleaved by proteases dysregulated in specific disease states. Upon peptide cleavage by a target protease, the barcodes are released from the ABN, after which they are small enough to diffuse into systemic circulation for subsequent renal clearance. By leveraging

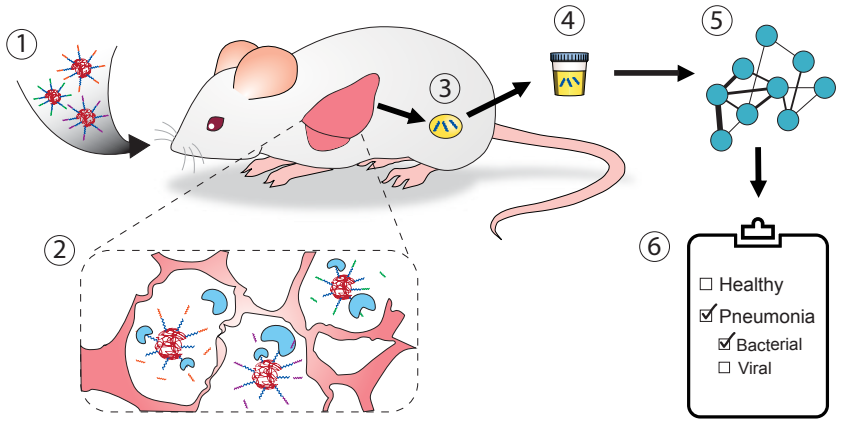


Figure 2-1: Schematic of approach. (1) ABNs are administered to mice that have been infected with either bacterial or viral pneumonia. (2) Proteases that are present in the lung cleave the ABNs at engineered substrate linkers, which releases the mass-encoded free reporters from the ABN scaffold into the circulation. (3) These reporters are filtered by the kidney and concentrated in the urine. (4) The reporters are then collected, and their concentrations are measured via mass spectrometry. (5) These concentrations are input into machine learning algorithms to train diagnostic classifiers. (6) This algorithm enables the diagnosis of pneumonia and in the case of infection, specifies whether the etiology is bacterial or viral.

catalytic protease activities and concentration of barcodes from a large circulating volume to a smaller urinary volume output, a highly amplified urinary readout is generated for sensitive disease detection. Furthermore, by multiplexing ABNs, we can create disease-specific urinary "signatures". Here, we have mined publicly available human transcriptomic datasets to create host-gene signatures for bacterial and viral pneumonia that consist solely of proteases. We then created a multiplexed panel of ABNs to detect the activity of a subset of these proteases, and demonstrated that they could produce protease-driven urinary signatures in mouse models of CAP that differ between bacterial and viral pneumonia within 2 hours of intrapulmonary sensor administration (Fig. 2-1).

2.2 Results

2.2.1 A 40-gene signature can distinguish bacterial and viral pneumonia

To create disease-specific gene signatures for bacterial versus viral pneumonia, we curated publicly available transcriptomic datasets for respiratory infections from whole blood (WB) and peripheral mononuclear blood cells (PBMCs), filtered these datasets for human proteases, and applied a computational multicohort framework designed to integrate gene expression data (MANATEE, Multicohort Analysis using AggregaTed gEne Expression) across 33 unique study cohorts (Table 2.1, Table 2.2, Fig. 2-2a).

Table 2.1: Discovery datasets for creating transcriptomic signatures. Cohorts of human transcriptomic data was used to train and validate a diagnostic classifier for bacterial and viral infections. Information on the source of the data and the patient cohorts that were used to train the classifier are listed.

Accession	Author	Platform	Tissue	Location	Demographic	Bacteria	Viruses	# of healthy samples	# of bacterial samples	# of viral samples
GSE101702	Yu	GPL21185	WB	Australia, Canada, Germany	Adults with influenza		Influenza	52	0	72
GSE117827	Tang	GPL23126	WB	USA	Children with acute viral infection		HRV, RSV, Enterovirus, Coxsackievirus	6	0	18
GSE16129 GPL96	Ardura	GPL96	PBMC	USA	Children with invasive staph infections	S. aureus		10	4	0
GSE17156	Zaas	GPL571	WB	USA, UK	Adults with respiratory viral infection		Influenza, HRV, RSV	56	0	27
GSE19491	Berry	GPL6947	WB	UK, South Africa	Patients with febrile bacterial infection	S. pyogenes, Staphylococcus spp.		18	75	0
GSE20346	Parnell	GPL6947	WB	Australia	Adults with CAP	Unknown	Influenza	18	6	4
GSE21802	Bermejo-Martin	GPL6102	WB	Spain	Adults with septic influenza		Influenza	4	0	12
GSE27131	Berdal	GPL6244	WB	Norway	Adults with influenza		Influenza	7	0	7
GSE38900 GPL10558	Mejias	GPL10558	WB	USA	Children with acute LRTI		RSV	8	0	28
GSE38900 GPL6884	Mejias	GPL6884	WB	USA, Finland	Children with acute LRTI		Influenza, HRV, RSV	31	0	153
GSE40012	Parnell	GPL6947	WB	Australia, Hong Kong	Adults with CAP	Unknown	Influenza	18	16	8
GSE42026	Herberg	GPL6947	WB	UK	Children admitted with febrile infections	Gram-positive	Influenza, RSV	33	18	41
GSE60244	Suarez	GPL10558	WB	USA	Adults hospitalized with LRTI	Unknown		40	22	0
GSE64456	Mahajan	GPL10558	WB	USA	Febrile infants ≤ 60 days old		Influenza, RSV, Enterovirus, HRV	19	0	108
GSE68310	Zhai	GPL10558	WB	USA	Adults with ARIs		Influenza, HRV, RSV, Enterovirus, Coronavirus	98	0	75
GSE82050	Tang	GPL21185	WB	Germany	Adults with influenza		Influenza	15	0	24

By applying a set of differential expression statistics and machine learning algorithms,

Table 2.2: Validation datasets of the transcriptomic signatures. Independent human datasets were used to test the diagnostic classifier. Information on the source of the data and the patient cohorts that were used to test the classifier are listed.

Accession	Author	Platform	Tissue	Location	Demographic	Bacteria	Viruses	# of healthy samples	# of bacterial samples	# of viral samples
E-MTAB-5195	Jong	GPL570	WB	Netherlands	Infants with RSV		RSV	4	0	39
GSE103842	Rodriguez-Fernandez	GPL10558	WB	USA	Young children hospitalized with bronchiolitis		RSV	12	0	62
GSE30119	Banchereau	GPL6947	WB	USA	Children with community acquired Staph infection	S. aureus		44	10	0
GSE34205	Ioannidis	GPL570	PBMC	USA	Children with ARIs		Influenza, RSV	22	0	79
GSE4607	Wong	GPL570	WB	USA	Septic children in the PICU	Multiple	Influenza	15	9	2
GSE6269 GPL96	Ramilo	GPL96	PBMC	USA	Children with bacterial or viral sepsis	S. aureus, S. pneumoniae	Influenza	6	12	3
GSE66099	Sweeney	GPL570	WB	USA	Septic children in the PICU	Multiple	Influenza, HMPV, Parainfluenza	47	35	5
GSE67059 GPL10558	Heinonen	GPL10558	WB	USA, Spain, Finland	Previously healthy children with asymptomatic or symptomatic HRV		HRV	16	0	20
GSE67059 GPL6947	Heinonen	GPL6947	WB	USA, Spain, Finland	Previously healthy children with asymptomatic or symptomatic HRV		HRV	21	0	80
GSE73072 (RSV DEE1)	Liu	GPL14604	WB	USA	Patients in the acute phase of a viral challenge study		RSV	20	0	9
GSE73072 (H3N2 DEE2)	Liu	GPL14605	WB	USA	Patients in the acute phase of a viral challenge study		Influenza	17	0	9
GSE73072 (H1N1 DEE3)	Liu	GPL14606	WB	USA	Patients in the acute phase of a viral challenge study		Influenza	22	0	9
GSE73072 (H1N1 DEE4)	Liu	GPL14607	WB	USA	Patients in the acute phase of a viral challenge study		Influenza	19	0	5
GSE73072 (H3N2 DEE5)	Liu	GPL14608	WB	USA	Patients in the acute phase of a viral challenge study		Influenza	21	0	8
GSE73072 (HRV UVA)	Liu	GPL14609	WB	USA	Patients in the acute phase of a viral challenge study		HRV	20	0	8
GSE73072 (HRV DUKE)	Liu	GPL14610	WB	USA	Patients in the acute phase of a viral challenge study		HRV	26	0	11
GSE77087	de Steenhuijsen Piters	GPL10558	WB	USA	Young children with mild and severe RSV disease		RSV	23	0	81

we identified a subset of 40 proteases that were consistently differentially expressed between bacterial and viral respiratory infections (Fig. 2-2b). With our previously described signature score model [59,69], the expression of these protease genes was able to distinguish bacterial and viral respiratory infections in 16 discovery cohorts (area under the receiver operating characteristics curve (AUROC) = 0.90, 95% confidence interval (CI): 0.841-0.959; Fig. 2-2c). Furthermore, in held-out, un-seen samples from the 16 discovery cohorts, data from the 40 proteases achieved an AUROC of 0.826 (95% CI: 0.706-0.946; Fig. 2-2c). Finally, in 17 completely independent cohorts, the expression pattern of these proteases maintained high accuracy in separating bacterial and viral respiratory infections (AUROC

= 0.905, 95% CI: 0.855-0.954; Fig. 2-2c).

The proteases used in these signatures were selected purely based on differential gene expression and classification power. To validate their biological relevance to acute respiratory infections, we input the respective signatures into a molecular functions database [70, 71]. Notably, the bacterial gene set was significantly associated with pathways that included neutrophil degranulation, innate immunity, and extracellular matrix (ECM) organization (Fig. 2-2d). Neutrophils are an integral component of the early innate immune response and have been known to play a significant role in the clearing of bacterial pneumonia via mechanisms that include bacterial killing, antimicrobial peptide (AMP) production, and recruitment of other innate immune cells [72]. The ECM has also been shown to influence bacterial adhesion and colonization, and is remodeled during tissue repair following inflammatory damage [73]. The viral protease set was significantly enriched in several different pathways, including apoptosis and deubiquitination (Fig. 2-2e). Apoptosis is a classic defense mechanism against viral infections [74], while deubiquitination has been implicated more recently in the antiviral response [75, 76]. The associations between these newly derived *in silico* signatures and known biological pathways gave us confidence that our 40-protease gene signature (ProSet) would prove to be active, functional players in pneumonia, and thus valuable biomarkers *in vivo*.

2.2.2 A pneumonia-specific panel of ABNs can be created based on proteases implicated in pneumonia and inflammation

Activity-based nanosensors produce a functional output for disease and possess an inherent amplification factor, in that single target enzymes trigger the release of multiple reporters. Therefore, rather than rely on endogenous gene expression, we sought to create ABNs that leverage the catalytic activity of our ProSet to generate a non-invasive and amplified urinary readout for distinguishing bacterial versus viral pneumonia. This goal necessitated identifying peptide substrates that are susceptible to cleavage by our ProSet enzymes for use as protease-cleavable linkers in our ABNs. From prior literature, we identified peptide

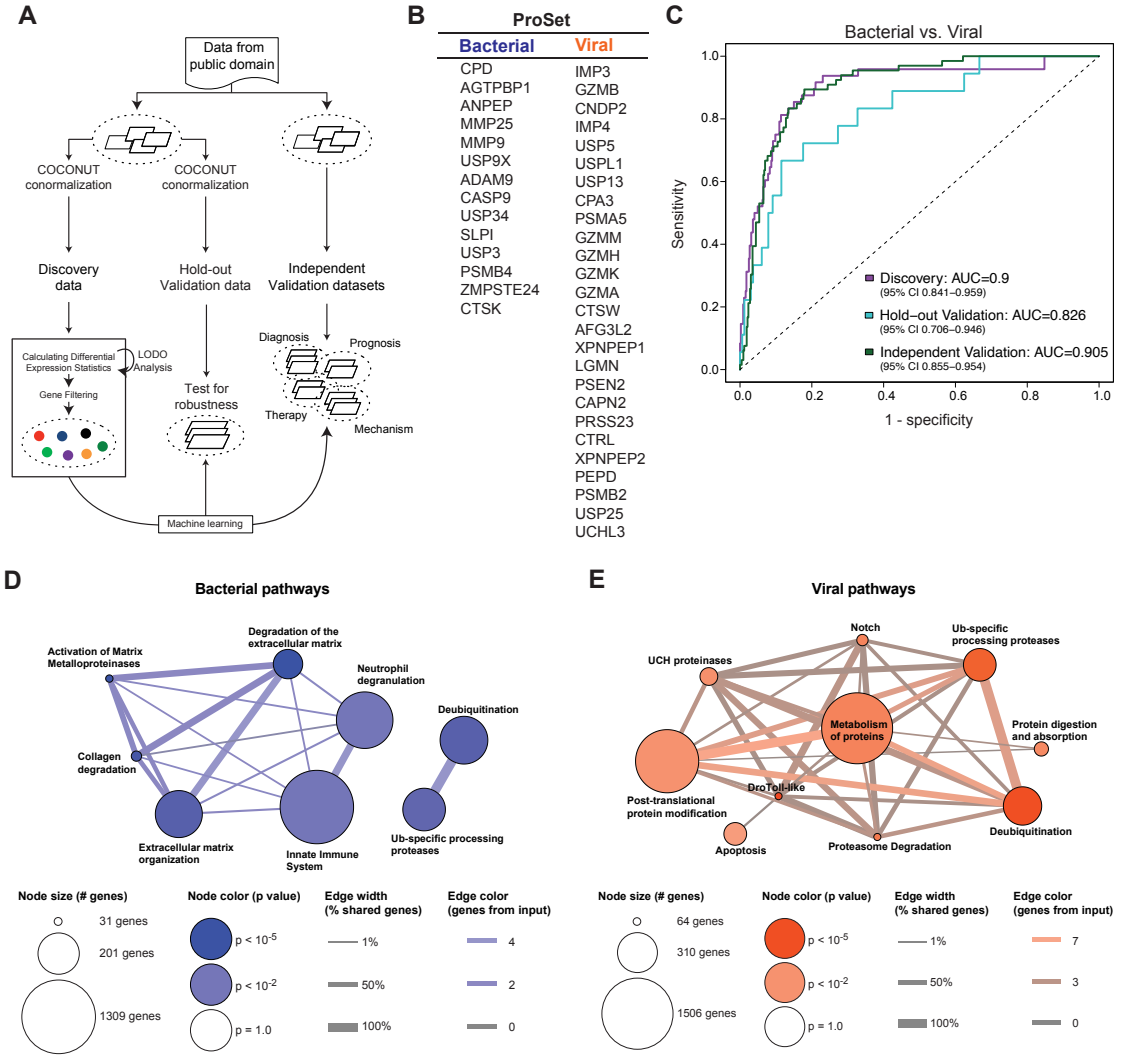


Figure 2-2: Generation of a bacterial versus viral infection protease signature using transcriptional metanalysis. (A) Publicly available transcriptional datasets from human patients with bacterial and viral respiratory infections were normalized using MANATEE, a computational framework for meta-analysis of gene expression data. (B) MANATEE yielded a 40-gene signature of proteases that are differentially upregulated in bacterial versus viral infections. (C) A classifier was trained on human data from 16 published cohorts and validated on 17 independent published cohorts. ROC curves represent the distinguishing power of the classifier, where an AUC of 0.5 indicates the classifier performs as well as chance and 1 indicates perfect classification. (D, E) Biological pathways underlying the different gene sets were queried using a pathway analysis program (ConsensusPathDB). The pathways are represented by nodes, with the size indicating the number of total genes associated with that pathway and the color indicating the significance of the inputted gene set in terms of its association with the pathway. Signature genes that are shared between different pathways are depicted as edges, with the color indicating the number of shared input genes.

substrates with amino acid sequences optimized for cleavage by a subset of the ProSet (see Table 2.3 for sequence source information). We then expanded the substrate pool by drawing upon our previous work, in which we identified peptide sequences that were efficiently cleaved in other mouse models of disease [55, 56, 64, 67], and nominated additional candidates based on literature review (references in Table 2.3). This process resulted in a panel of 20 peptides, which were incorporated into ABNs (Table 2.3). These ABNs consist of a 40-kDa polyethylene glycol (PEG) core that is conjugated to the peptide substrates. The N-terminus of each substrate linker is synthesized with a mass-encoded GluFib reporter, a peptide that is stable in circulation and renally cleared, to enable detection in urine via mass spectrometry. However, the N-terminus can also accommodate other reporter molecules, such as a fluorophore or short peptide domain, to suit different readout modalities.

Table 2.3: 20-plex panel of activity-based nanosensors for pneumonia. Each nanoparticle consists of a mass-encoded GluFib reporter (in italics, d-amino acids represented by lowercase letters) attached to a photolabile linker (ANP). The adjacent peptide sequence (in bold) is conjugated to an inert 40kDa 8-arm PEG scaffold. All peptide sequences were derived from published studies (reference listed per substrate).

Name	Sequence (*reporter*, **peptide sequence**)	Sequence source
BV01	*e(+2G)(+6V)ndneeGFFsAr*-(ANP)-**GGAIEFDSGC**-(PEG8-40kDa)	Published Granzyme B substrate (Kwong et al., 2020)
BV02	*eG(+6V)ndneeGF(+1F)s(+1A)r*-(ANP)-**GGHPGGPQC**-(PEG8-40kDa)	Commercially available Cathepsin K substrate
BV03	*e(+3G)(+1V)ndneeGFFs(+4A)r*-(ANP)-**GGGVFRMLSVGC**-(PEG8-40kDa)	Screened Granzyme A substrate (Kaiserman et al., 2014)
BV04	*e(+2G)Vndnee(+2G)FFs(+4A)r*-(ANP)-**GGGLFRSLSSGC**-(PEG8-40kDa)	Screened Granzyme A substrate (Kaiserman et al., 2014)
BV05	*eGVndnee(+3G)(+1F)Fs(+4A)r*-(ANP)-**GGGLLYGKGGC**-(PEG8-40kDa)	Published CAPN2 substrate (Kwon et al., 2020)
BV06	*e(+2G)(+6V)ndnee(+3G)(+1F)(+1F)s(+1A)r*-(ANP)-**GGy-Tic-TNGC**-(PEG8-40kDa)	Published Legumain substrate (Poreba et al., 2018)
BV07	*eG(+6V)ndnee(+3G)(+1F)Fs(+4A)r*-(ANP)-**GGfPRSGGGC**-(PEG8-40kDa)	Derived from Kirkpatrick et al., 2020
BV08	*e(+3G)(+1V)ndneeG(+10F)FsAr*-(ANP)-**GGGSGRSANAKGC**-(PEG8-40kDa)	Derived from Kirkpatrick et al., 2020
BV09	*e(+2G)Vndnee(+2G)F(+10F)sAr*-(ANP)-**GGGIQQRSLGGGC**-(PEG8-40kDa)	Derived from Dudani et al., 2018
BV10	*eGVndneeGF(+10F)s(+4A)r*-(ANP)-**GGIPSIQSRGLGC**-(PEG8-40kDa)	Influenza hemaagluttinin cleavage site (Tse and Whittaker, 2015)
BV11	*e(+2G)(+6V)ndneeG(+10F)(+1F)s(+1A)r*-(ANP)-**GGNLARALKQTIGC**-(PEG8-40kDa)	Screened MMP substrate (Kukreja et al., 2015)
BV12	*eG(+6V)ndneeG(+10F)Fs(+4A)r*-(ANP)-**GGHMVQHLIQWHGC**-(PEG8-40kDa)	Screened MMP substrate (Kukreja et al., 2015)
BV13	*e(+3G)(+1V)ndnee(+2G)(+10F)Fs(+4A)r*-(ANP)-**GGPRAAA-Homophe-TSPGC**-(PEG8-40kDa)	Screened ADAM9 substrate (Moss et al., 2016)
BV14	*e(+2G)Vndnee(+3G)(+10F)(+1F)s(+4A)r*-(ANP)-**GGTGPPGYTGC**-(PEG8-40kDa)	Screened ADAMTS substrate (Bekhouche et al., 2016)
BV15	*eGVndneeG(+10F)(+10F)sAr*-(ANP)-**GGTGLPVYQGC**-(PEG8-40kDa)	Screened ADAMTS substrate (Bekhouche et al., 2016)
BV16	*e(+2G)(+6V)ndnee(+3G)(+10F)(+1F)s(+4A)r*-(ANP)-**GG-Nle(O-Bzl)-Met(O)2-Oic-Abu-C**-(PEG8-40kDa)	Published Neutrophil elastase substrate (Kasperkiewicz et al., 2014)
BV17	*eG(+6V)ndneeG(+10F)(+10F)sAr*-(ANP)-**GGAAFAGC**-(PEG8-40kDa)	Published Neutrophil elastase substrate (Buss et al., 2018)
BV18	*e(+3G)(+1V)ndnee(+2G)(+10F)(+10F)sAr*-(ANP)-**GGGGGPGC**-(PEG8-40kDa)	Derived from Dudani et al., 2018
BV19	*e(+2G)VndneeG(+10F)(+10F)s(+4A)r*-(ANP)-**GGPLGMRGGC**-(PEG8-40kDa)	Derived from Kirkpatrick et al., 2020
BV20	*eGVndnee(+2G)(+10F)(+10F)s(+4A)r*-ANP-**GGP-(Cha)-G-Cys(Me)-HAGC**-(PEG8-40kDa)	Derived from Kirkpatrick et al., 2020

2.2.3 The pneumonia ABN panel generates etiology-specific urinary signatures

We next sought to determine whether this ABN panel could distinguish bacterial from viral pneumonia *in vivo* by administering the ABN panel into five mouse models representing common causes of CAP in humans. We established the models by infecting immunocompetent BALB/c mice with either bacteria (*Streptococcus pneumoniae*, SP; *Klebsiella pneumoniae*, KP; *Haemophilus influenzae*, HI) or viruses (Influenza A/PR/8/34 (H1N1), PR8; Pneumonia virus of mice, PVM). We optimized the dose of each pathogen to cause similar timelines of disease within each etiology based on lung bacterial and viral loads, and outward physical symptoms (Figure 2-3).

To characterize the performance of our panel *in vivo*, we then delivered the 20 ABNs directly into the lungs of mice with bacterial pneumonia, viral pneumonia, or healthy controls, and collected urine two hours after administration. The urinary reporter concentrations were measured, normalized, and compared to assess whether the ABNs were differentially cleaved among disease states. Principal component analysis (PCA) revealed distinct divergence between mice with pneumonia and the corresponding healthy controls, consistent with the hypothesis that ABNs are differentially cleaved between these groups (Fig. 2-4b). Furthermore, the tightness of each group indicates that these cleavage events are consistent, suggesting the existence of ABN-generated urinary signatures for mice with and without pneumonia. To determine these signatures, we examined the relative differences in reporter concentrations between the healthy and infected mice and found that 17 of the 20 ABNs were significantly differentially cleaved in one state versus the other (Fig. 2-4c). This demonstrated that the host protease response could be queried to generate a functional readout of active pneumonia. Furthermore, it is worth noting that even though the infected lung is presumably more protease-rich than healthy tissue, many reporters are significantly enriched in healthy controls relative to the mice with pneumonia, which highlights differential protease activity between the two states.

Re-labeling each mouse by etiology revealed further separation between bacterial versus

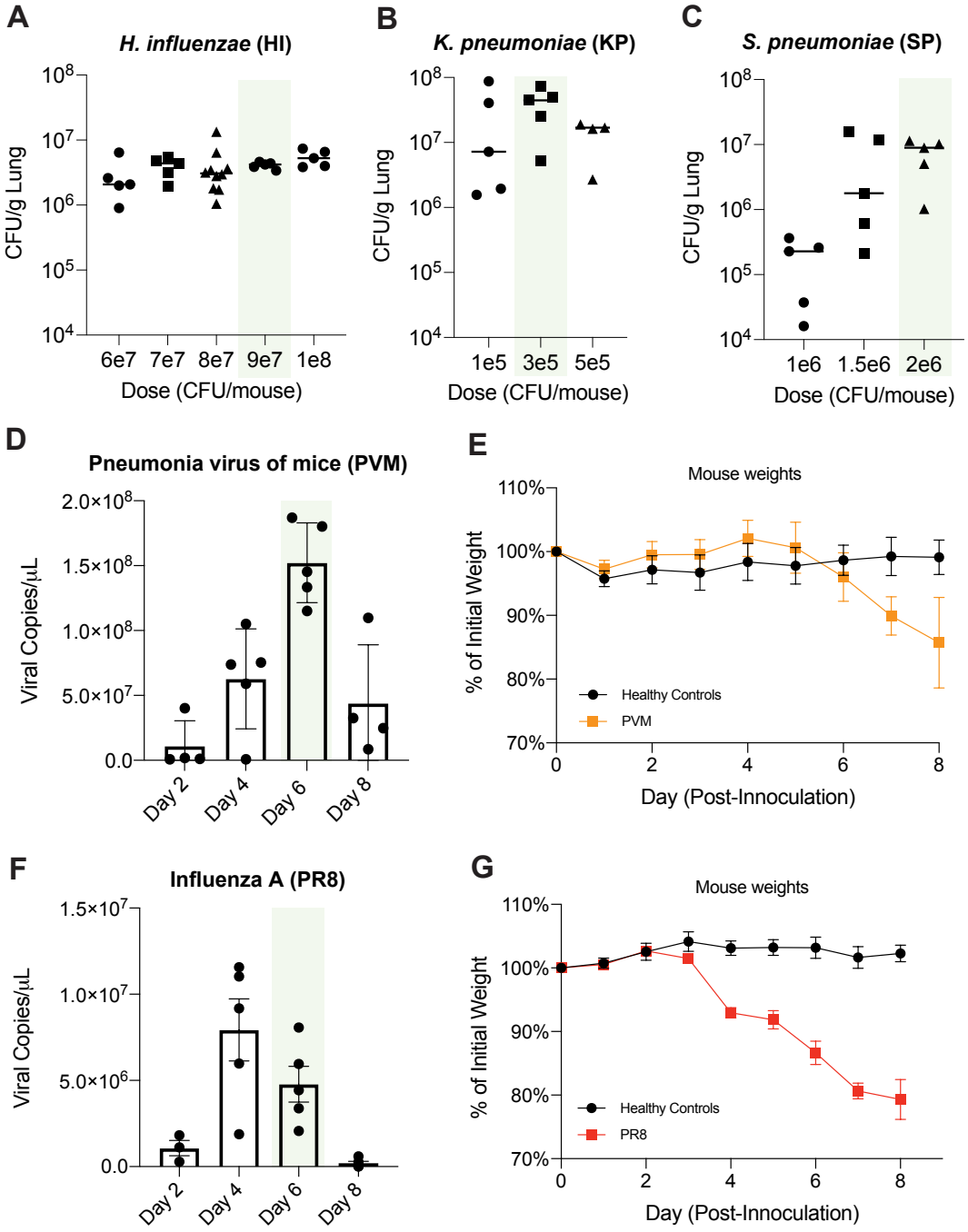


Figure 2-3: Characterization of the mouse models for bacterial and viral pneumonia. (A,B,C) Various doses of each bacteria were administered to immunocompetent mice. Lungs from these mice were homogenized and plated to determine bacterial loads. Each point represents one mouse, n = 5 to 10 per dose. (D,F) The viral load in mice infected with with pnuemonia virus of mice (PVM) and influenza A (PR8) was evaluated over time. Viral loads were quantified using qRT-PCR. n=3-5 mice per timepoint. (E,G) The physical manifestations of disease were tracked via body weight throughout the timecourse of infection. n=3-5 mice per timepoint.

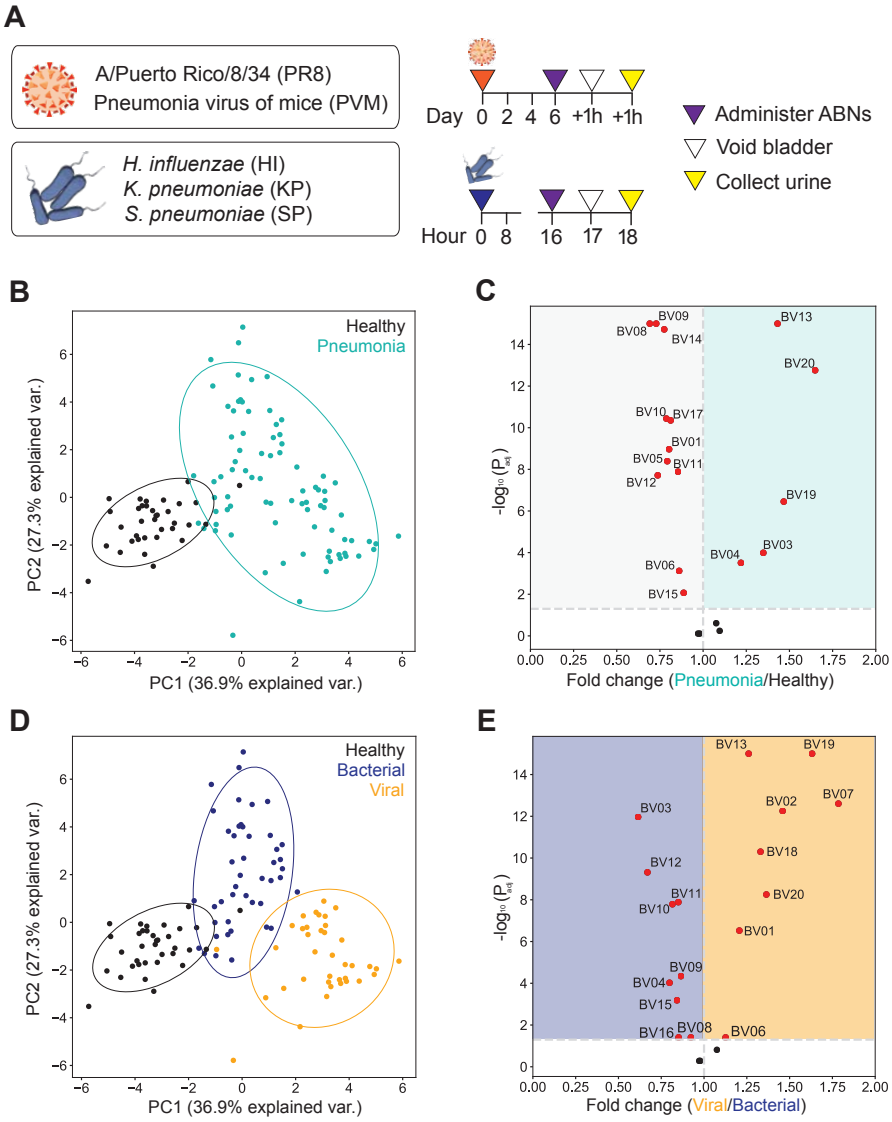


Figure 2-4: Activity-based nanosensors distinguish pneumonia and etiology in mice. (A) ABNs were administered into 5 mouse models of pneumonia. Urine from each mouse was collected 2 hours after administration to characterize *in vivo* ABN activity. (B,D) Unsupervised principal component analysis (PCA) of normalized urine reporter concentrations in pneumonia (n=106 mice) and healthy controls (n=35 mice). Data from pneumonia mice are labeled according to either infection (B) or etiology (D). (C,E) The relative fold change between disease states was calculated using mean scaled reporter concentrations. Dotted line at x-axis represents no fold change between disease states. Each point represents one reporter, with significantly differential reporters in red. Significance was calculated using two-tailed t-test with Holm-Sidak correction. Dotted y-axis at P_{adj} = 0.05.

viral pneumonia (Fig. 2-4d). Notably, of the 5 reporters enriched in mice with pneumonia compared to controls, 3 (BV13, BV19, BV20) were significantly enriched in the viral mice while the other 2 (BV03 and BV04) were enriched in the bacterial mice (Fig. 2-4e). Reporters from additional ABNs that were not part of this initial subset (e.g., BV01, BV10, BV12) also emerged as differentially enriched in bacterial and viral pneumonia. The association of these reporters with the healthy controls when comparing to mice infected with pneumonia, versus bacterial or viral pneumonia when comparing etiologies, demonstrates that multiplexing ABNs enables the creation of discrete reporter sets for different disease states. Overall, these results revealed distinct differences in urinary reporter concentrations, and thus protease activity, between mice with bacterial and viral pneumonia.

2.2.4 Activity-based nanosensors are cleaved by a wide range of protease classes

Having shown that the pneumonia activity-based nanosensor panel generated differential *in vivo* cleavage signatures, we next sought to evaluate the specificity of each substrate for its target protease by screening of the peptide linkers across a non-exhaustive range of proteases. To this end, we first reformulated each ABN into a fluorescent probe format by flanking the peptide substrate sequence with a fluorophore-quencher pair (Fig. 2-5a; sequences listed in Table 2.4). We then incubated each probe with commercially-available recombinant proteases that were either derived from the ProSet or predicted to be present in the lungs (protease and buffer conditions in Table 2.5).

All probes were cleaved by at least one protease within 10 minutes, with marked cleavage by several proteases that are known to be more promiscuous, such as neutrophil elastase (NE) and matriptase (ST14) (Fig. 2-5a). We performed hierarchical clustering to determine which proteases had orthogonal cleavage profiles and saw similarities across Cathepsin K (CATK), Kallikrein 5 (KLK5) and Serine Protease 3 (PRSS3). However, we observed no obvious clustering based on protease class, and in fact, most proteases had relatively distinct cleavage patterns, which was expected given our rational approach to substrate selection.

Table 2.4: Quenched fluorescent probe formulations of ABN panel. Substrate sequences were incorporated into quenched fluorescent probes. 5FAM = Fluorescein, CPQ2 = Quencher, PEG2 = polyethylene glycol. Other capital letters are single amino acid codes.

Substrate	Sequence
BV01-F	(5FAM)-GGAIEFDSGK(CPQ2)-(PEG2)-C
BV02-F	(5FAM)-GGHPGGPQGK(CPQ2)-(PEG2)-C
BV03-F	(5FAM)-GGGVFRMLSVGK(CPQ2)-(PEG2)-C
BV04-F	(5FAM)-GGGLFRSLSSGK(CPQ2)-(PEG2)-C
BV05-F	(5FAM)-GGGLLYGKGGK(CPQ2)-(PEG2)-C
BV06-F	(5FAM)-GGy-Tic-TNGK(CPQ2)-(PEG2)-C
BV07-F	(5FAM)-GGfPRSGGGK(CPQ2)-(PEG2)-C
BV08-F	(5FAM)-GGGSGRSANAKGK(CPQ2)-(PEG2)-C
BV09-F	(5FAM)-GGGIQQRSLGGGK(CPQ2)-(PEG2)-C
BV10-F	(5FAM)-GGIPSIQSRGLGK(CPQ2)-(PEG2)-C
BV11-F	(5FAM)-GGNLARALKQTIGK(CPQ2)-(PEG2)-C
BV12-F	(5FAM)-GGHMVQHLIQWHGK(CPQ2)-(PEG2)-C
BV13-F	(5FAM)-GGPRAAA-Homophe-TSPGK(CPQ2)-(PEG2)-C
BV14-F	(5FAM)-GGTGPPGYTGK(CPQ2)-(PEG2)-C
BV15-F	(5FAM)-GGTGLPVYQGK(CPQ2)-(PEG2)-C
BV16-F	(5FAM)-GG-Nle(O-Bzl)-Met(O)2-Oic-Abu-K(CPQ2)-(PEG2)-C
BV17-F	(5FAM)-GGAAFAGK(CPQ2)-(PEG2)-C
BV18-F	(5FAM)-GGGGGPGK(CPQ2)-(PEG2)-C
BV19-F	(5FAM)-GGPLGMRGGK(CPQ2)-(PEG2)-C
BV20-F	(5FAM)-GGP-(Cha)-G-Cys(Me)-HAGK(CPQ2)-(PEG2)-C

Table 2.5: Recombinant proteases and buffers used for *in vitro* screen. Specific buffers were used to create optimal cleavage conditions for each recombinant protease. Activation buffers were used for pre-incubation of the protease as needed.

Protease	Product info	Assay buffer	Activation buffer
rhFAP	R&D (3715-SE)	50 mM Tris, 1 M NaCl, 1 mg/mL BSA, pH 7.5	
rhNE	Enzo BML-SE284-0100	0.1 M Tris-HCl pH 8	
rhLGMN	R&D (2199-CY)	50 mM MES, 250 mM NaCl, pH 5.0	50 mM Sodium Acetate, 100 mM NaCl, pH 4.0
rhMMP9	Enzo (BML-SE360)	50 mM Tris, 0.15 M NaCl, 10 mM CaCl2, 0.05% Brij-35 (w/v), pH 7.5	
rhPR3	Enzo (BML-SE498-0025)	100mM MOPS pH 7.5, 500mM NaCl, 10% DMSO, 100µM DTNB	
rhMMP24	Enzo (ALX-201-105-C010)	50mM TRIS-HCl, pH 7.57, 150mM NaCl, 5mM CaCl2, 0.025% Brij 35.	
rhGZMK	Enzo (ALX-201-117-C010)	50 mM TRIS, pH 8.0, 0.15M NaCl, 0.01% Triton X-100, 0.3mM DTNB	
rhGZMA	R&D (2905-SE)	50 mM Tris, pH 8.0	0.1 M Tris, pH 9.0
r3CLpro	R&D (E-720)	50 mM HEPES, 0.1 M NaCl, pH 8	
rhCATB	R&D (953-CY-010)	25 mM MES, pH 5.0	25 mM MES, 5 mM DTT, pH 5.0
rhNAPSA	R&D (8489-NA-050)	50 mM Sodium Acetate, 100 mM NaCl, pH 4.0	
rhGZMH	R&D (1377-SE-010)	50 mM Tris, 1 M NaCl, 1 mg/mL BSA, pH 7.5	
rhCATG	Enzo (BML-SE283-0100)	160 mM Tris-HCl, 1.6 M NaCl, pH 7.7	
rPLPro	R&D (E-611)	50 mM HEPES, 0.1 M NaCl, pH 8	
rhADAM9	R&D (939-AD-020)	25 mM Tris, 2.5 µM ZnCl2, 0.005% (w/v) Brij-35, pH 9.0	
rhGZMB	R&D (2906-SE)	50 mM Tris, pH 7.5	50 mM MES, 50 mM NaCl, pH 5.5
rhST14	R&D (3946-SEB-010)	50 mM Tris, 0.05% (w/v) Brij-35, pH 9.5	
rhF2	R&D (1473-SE-010)	50 mM Tris, 1 M NaCl, 1 mg/mL BSA, pH 7.5	
rhCATK	Enzo (BML-SE553-0010)	25 mM MES, 5 mM DTT, pH 5.0	
rhKLK5	R&D (1108-SE-010)	0.1 M NaH2PO4, pH 8.0	
rhPRSS3	R&D (3714-SE)	50 mM Tris, 0.15 M NaCl, 10 mM CaCl2, 0.05% Brij-35 (w/v), pH 7.5	

Importantly, although the peptide substrates were selected based on published susceptibility to cleavage by our proteases of interest, peptide sequences are often cleaved by multiple enzymes, to varying degrees; thus, there will inevitably be substrate cleavage by other proteases *in vivo*, especially in the protease-rich microenvironment of the infected lung. To parse through this confounding cleavage and more directly correlate our *in vivo* results with specific protease activity, we normalized the screening data to identify the strongest protease-substrate pairs from our screen. We reasoned that standardizing across substrates would allow us to compare how quickly an individual protease cleaved one probe relative to others, thus providing a yardstick for cleavage efficiency. Conversely, normalizing across proteases reflected specificity by comparing the cleavage rates of one probe by a wide panel of proteases. Correlating these metrics enabled the identification of optimal protease-substrate pairs that had both robust and specific cleavage (Fig. 2-5b, Fig. 2-6). Based on their cleavage efficiency and specificity, several optimal protease-probe pairs emerged from the screen. Pairings such as Granzyme B (GZMB) and BV01-F (Fig. 2-5b), and Legumain (LGMN) and BV06-F (Fig. 2-5c) confirmed that some of the rationally designed probes were being well cleaved by their intended targets (as listed in Table 2.3). Other probes yielded no optimal protease hits, though it is possible that an unscreened protease might cleave that probe more successfully, or simply that the optimal cleavage kinetics for that probe is not achieved during a 10-minute assay. Overall, this analysis indicated that while each ABN is vulnerable to protease cleavage, the presence of any given reporter in the urine may not directly correspond to one specific protease. Still, given that the cleavage patterns *in vitro* were sufficiently distinct, and that multiplexing allows us to leverage the relative reporter levels from each ABN, we anticipated that our *in vivo* cleavage signatures would enable classification among disease states.

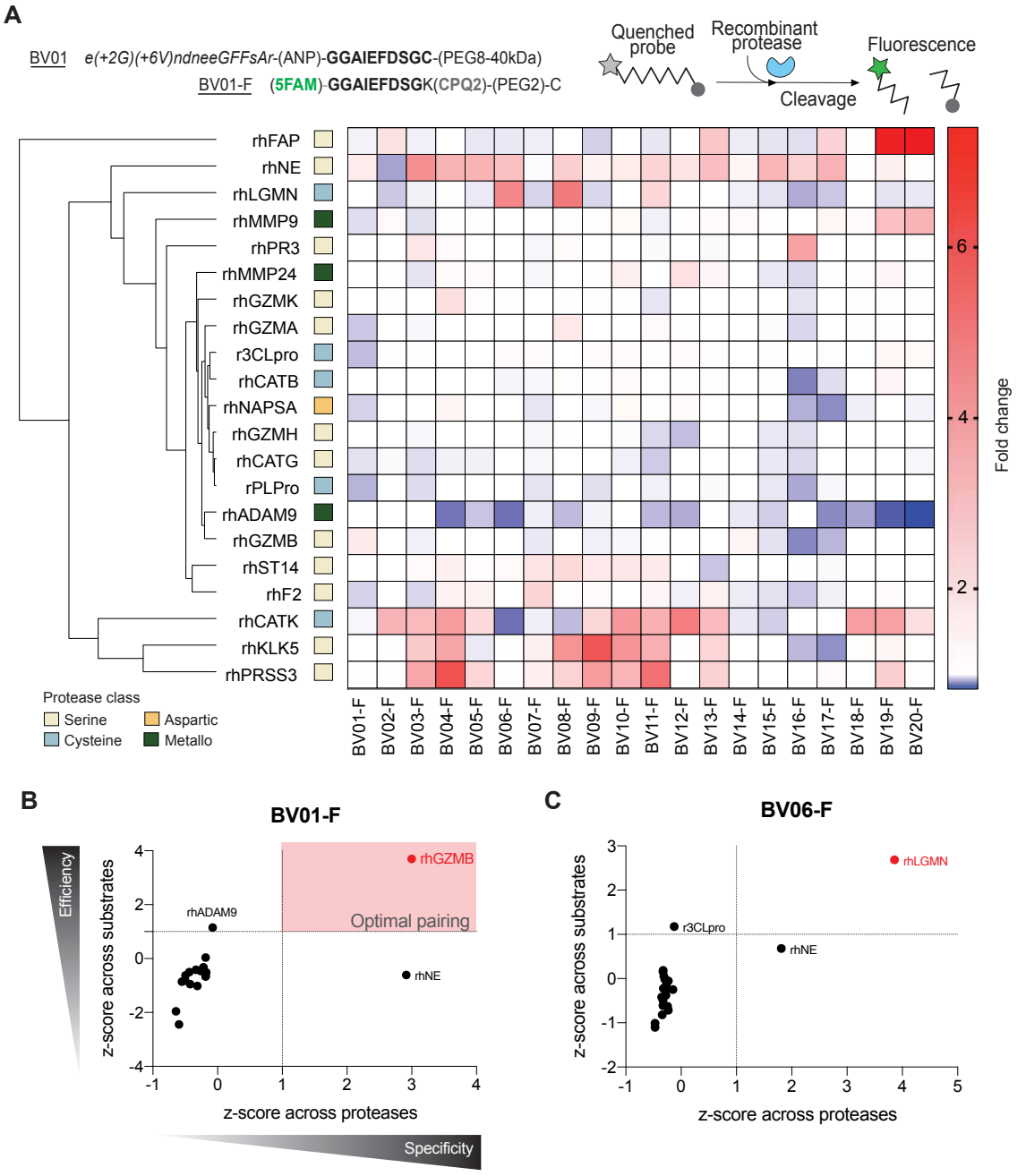


Figure 2-5: In vitro screening of fluorescent substrates reveals possible ABN targets. (A) Each peptide sequence was also incorporated into a quenched fluorescent substrate. These fluorogenic probes were then incubated with recombinant proteases to evaluate the cleavage profile of each ABN. Hierarchical clustering was performed based on the fold change in fluorescence after 10 minutes (average of two replicates). (B,C) Standardization was performed to assess protease-substrate pairings from the in vitro screening data. Z-scores of the average fold change values for each pairing across the proteases (x-axis) and substrates (y-axis) were compared to characterize protease-substrate pairs by specificity and cleavage efficiency.

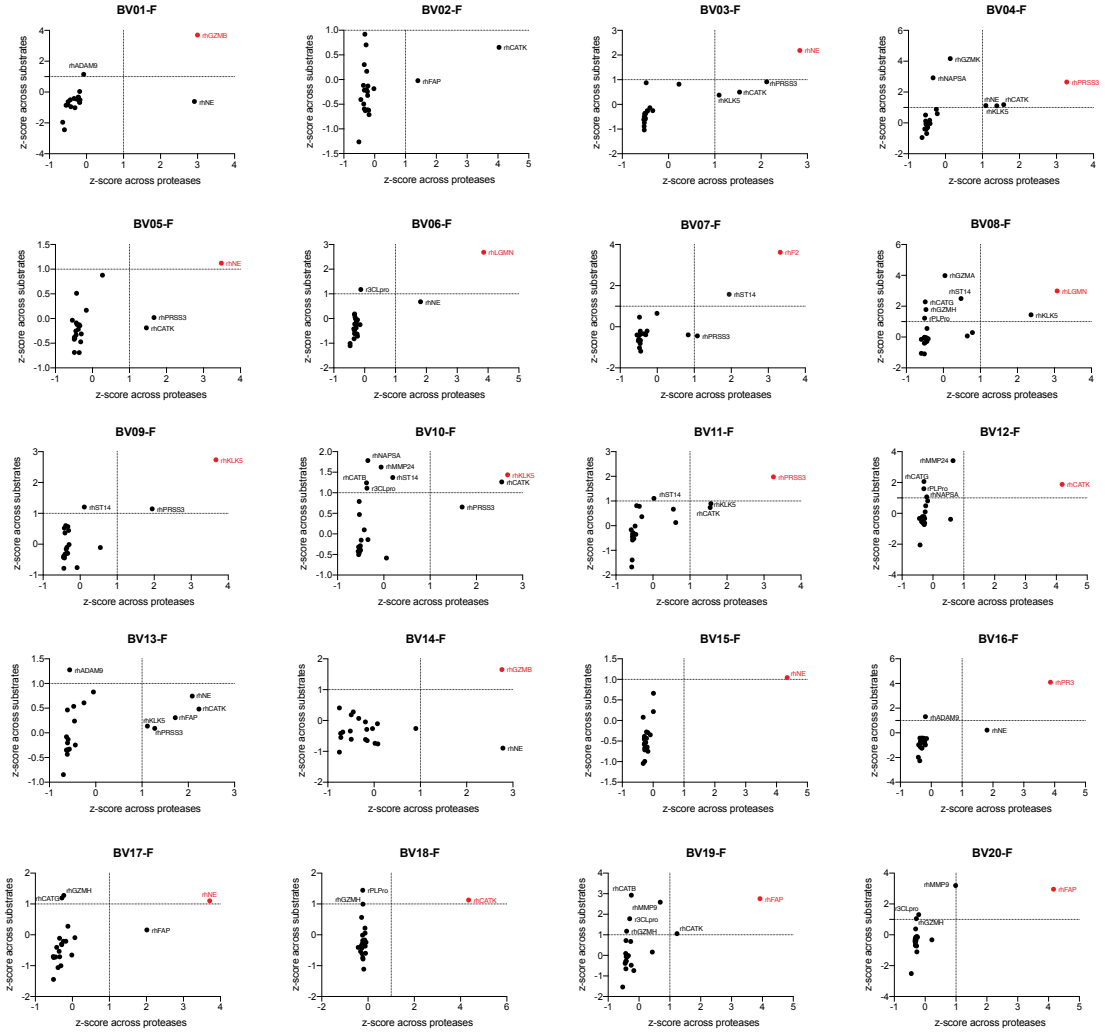


Figure 2-6: Z-score correlation plots with "optimal" proteases. Each plot represents standardized metrics that were calculated based on the fluorescence fold change at 10 minutes after incubation of the fluorescent probe with each recombinant protease. The most protease with the highest of both metrics is in red, but other proteases in the upper red quadrant are also considered "optimal" hits.

2.2.5 BV01 signals differences in the host immune response to bacterial and viral pneumonia

After observing differential cleavage of the ABNs *in vivo* and characterizing cleavage susceptibilities *in vitro*, we sought to validate that our urine reporter signatures were reflective of diverging biology between bacterial and viral pneumonia. To do this, we focused on BV01, which was designed and validated to be cleaved by Granzyme B (GzmB) *in vitro*. GzmB is a serine protease that is produced by natural killer (NK) cells and cytotoxic T lymphocytes (CTLs), and has been implicated in the antiviral response. [77] NK cells in particular are known to play an important role in antiviral immunity. [78] However, there is evidence that they have a detrimental effect on the lungs of immunocompromised mice in the context of *S. pneumoniae* due to the production of inflammatory cytokines. [79] To determine whether our differential BV01 signal was being driven by the immune response to viral infection, we focused on the main causes of viral and bacterial pneumonia: influenza (PR8) and *S. pneumoniae* (SP), respectively. We hypothesized that the virally-associated BV01 signal we observed *in vivo* (Fig. 2-4e) was due to increased recruitment of activated NK cells and CTLs into the lungs during viral infection. To assess the presence of these immune cells in viral and bacterial pneumonia, we performed immunofluorescent staining for cell surface markers on CTLs (CD8 co-receptor) and NK cells (NKp.46) using fresh frozen lung sections from SP and PR8 mice. The staining showed few CD8+ T cells and NK cells in the SP tissue (Fig. 2-7a,b,f) relative to the PR8-infected lungs, which had a higher number of positive cells for both cell populations. To confirm that the low levels of staining in the SP samples were due to negative signal rather than incomparable infection or a lack of tissue binding, we also stained for an antibody against RB6-8C5, which reacts strongly with Ly6G, a neutrophil marker. We observed robust staining of RB6-8C5+ cells in SP compared to PR8 (Fig. 2-7c,h), which contrasted with the CD8 and NK staining patterns and indicated that the SP lungs were indeed inflamed. We also stained for GzmB protein in the same sections and saw increased staining in PR8-infected lungs relative to SP-infected samples (Fig. 2-7d,g), which aligned with elevated GzmB expression in PR8 lungs, as mea-

sured by qRT-PCR (Fig. 2-7e). Overall, these results show that GzmB, CD8+ T cells and NKp46+ NK cells are significantly more prevalent in PR8 lungs compared to SP, which is consistent with previous findings that NK cells are actively recruited to the lungs following influenza infection in mice. [80]

At this point, we had nominated GzmB as a viral target via transcriptomics, shown that BV01 is efficiently and specifically cleaved by GzmB *in vitro*, observed staining that suggests that the cells producing GzmB are enriched in the deep lung infected with PR8 compared to SP, and validated that GzmB itself is upregulated in PR8 based on immunohistochemistry and mRNA expression. We next sought to test whether the BV01 urinary signal could be considered a proxy for GzmB activity in the context of pneumonia. To do this, we constructed an activatable zymography probe (AZP) using the BV01 substrate (BV01-Z) to observe *in situ* cleavage in fresh frozen lung tissue samples from our mouse models of infection (Fig. 2-8a). AZPs are composed of an anionic poly-glutamic acid (polyE) domain that is connected to a fluorophore-labeled cationic poly-arginine (polyR) domain via a peptide substrate. [81] When the peptide substrate adjoining the polyE and polyR domains in the AZP (in this case, the BV01 linker) is cleaved by a protease, the AZP is activated and the polyE and polyR domains can separate. Freed cationic polyR domains can then bind locally to the tissue where the AZP was activated, enabling *in situ* labeling of protease activity on fresh frozen tissue sections (Fig. 2-8a). To confirm that BV01-Z could be cleaved by GzmB and that proteolytic activation was necessary for tissue binding, we incubated the AZP with recombinant GzmB, allowing protease-driven activation to take place. We then applied either the pre-cleaved mixture or intact probe onto fresh frozen sections of healthy mouse lung at 4ºC, enabling tissue binding while preventing activity of endogenous enzymes. A fluorescent signal was visible in the sections incubated with the pre-cleaved AZP, but not with the intact probe, demonstrating that a positive signal is dependent on protease activity (Fig. 2-8b).

Having validated that GzmB cleaved BV01-Z, we applied BV01-Z to fresh frozen lung sections from PR8 infected mice and observed strong signal throughout the lungs (Fig. 2-

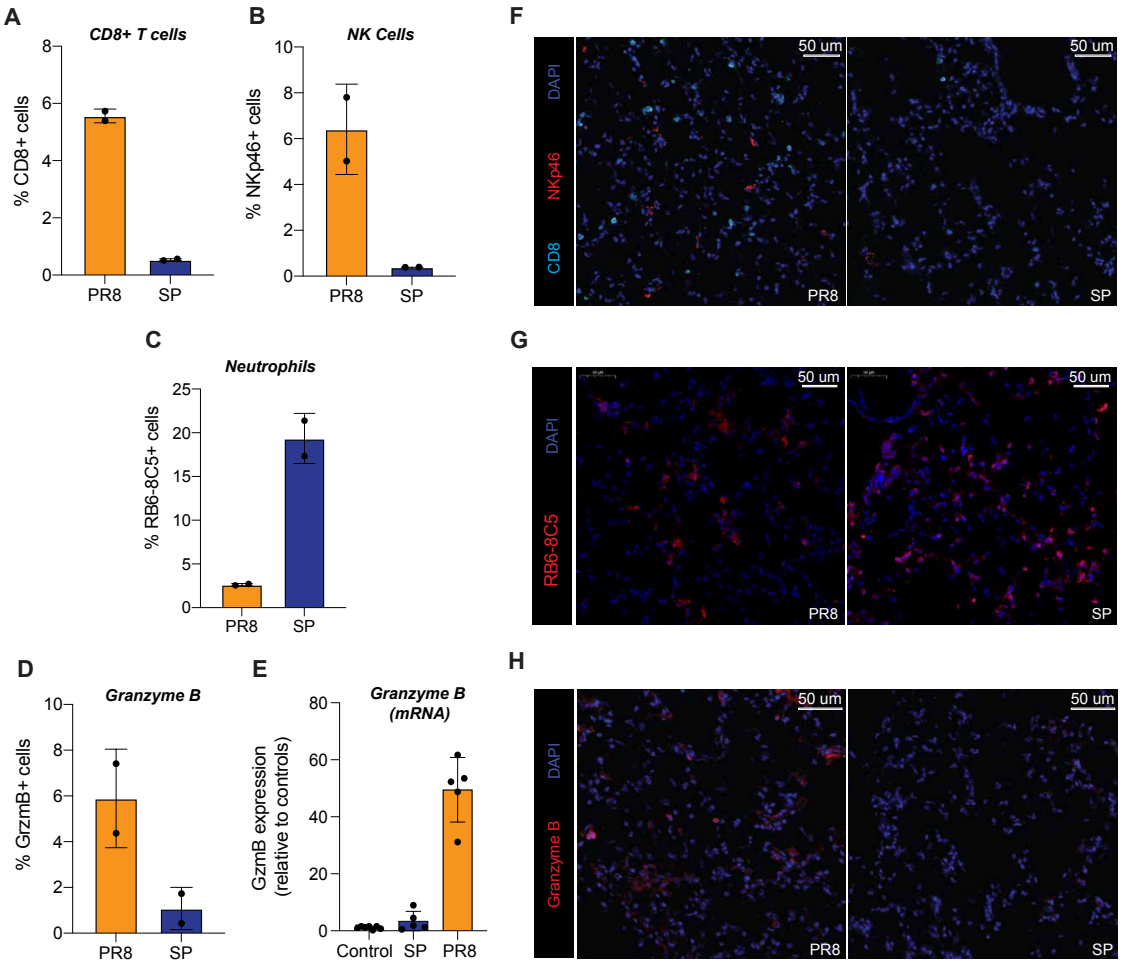


Figure 2-7: Granzyme B is elevated in influenza pneumonia. (A-D) Percentage of detected cells positive for various cell markers via immunofluorescent staining performed on fresh frozen sections from mice infected with either influenza (PR8) or S. pneumoniae (SP) (n=2 consecutive sections per group, mean +/- s.d.). Counts for positive/total cells per section in each panel are (A,F) CD8 (PR8: 4323/80149, 4765/83153; SP: 441/86308, 716/127753) (B,F) NKp46 (PR8: 4022/80149, 6488/83153; SP: 339/86308, 500/127753), (C,G) RB6-8C5 (PR8: 2650/97205, 2298/90169; SP: 22351/129038, 30945/144752), and (D,H) GzmB (PR8: 7019/94643, 3642/83318; SP: 1565/90409, 519/121635). (F-H) Representative images of influenza (PR8) and S. pneumoniae (SP) lungs stained for (F) natural killer cells (NKp.46) and CD8 T-cells (CD8), (G) neutrophils (RB6-8C5) and (H) Granzyme B. Scale bars = 50 μm. Counterstain with DAPI (blue).(E) Relative expression of GzmB in lung tissue from healthy control mice and those with PR8 or SP via qRT-PCR.

8c). This labeling was significantly abrogated by the addition of a GzmB specific inhibitor (Z-AAD-CH2Cl) (Fig. 2-8c, p=0.0180), confirming that *in situ* GzmB activity contributes to BV01-Z activation. However, the signal is not completely eliminated with the inhibitor, indicating some off-target cleavage by other proteases present in the tissue. We confirmed this orthogonal cleavage by incubating BV01-Z with a broad-spectrum protease inhibitor cocktail, and observed a trend in further reduction of BV01-Z signal relative to GzmB specific inhibition, but it was not significant (Fig. 2-8c, n.s. p=0.0563). Nevertheless, a certain level of off-target cleavage was to be expected as the *in vitro* screen revealed that BV01-F is also cleaved to a notable, but lesser, extent by NE and ADAM9 (Fig. 2-5b), and is also likely cleaved by other proteases that were not included in the recombinant screen. We also compared BV01-Z signal in PR8 versus healthy lung, and observed increased BV01-Z signal in the infected tissue (Fig. 2-8d). Taken together, we have shown that there is a greater influx of natural killer and CD8+ T cells in viral pneumonia compared to bacterial pneumonia. These cells produce the viral pneumonia marker GzmB, which can cleave the BV01 peptide sequence to produce a reporter signal that is appreciable in urine samples and tissue sections from mice with viral pneumonia.

2.3 Discussion

Our results reveal several new insights into diagnosing community-acquired pneumonia (CAP). First, we demonstrated that proteases can be mined from existing transcriptomic human data and leveraged as novel biomarkers for distinguishing pneumonia etiology. Existing biomarkers for pneumonia such as as C-reactive protein (CRP) and procalcitonin (PCT) require measurement from blood. Similarly, our *in silico* protease signature was derived from transriptomic data derived from whole blood and PBMCs. However, the diagnostic value of CRP and PCT is dependent on their concentrations in blood, which is diluted by the circulatory volume and accumulates from the whole body. In contrast, we have previously shown that a vast majority of ABNs that are intratracheally delivered directly into the lungs remain

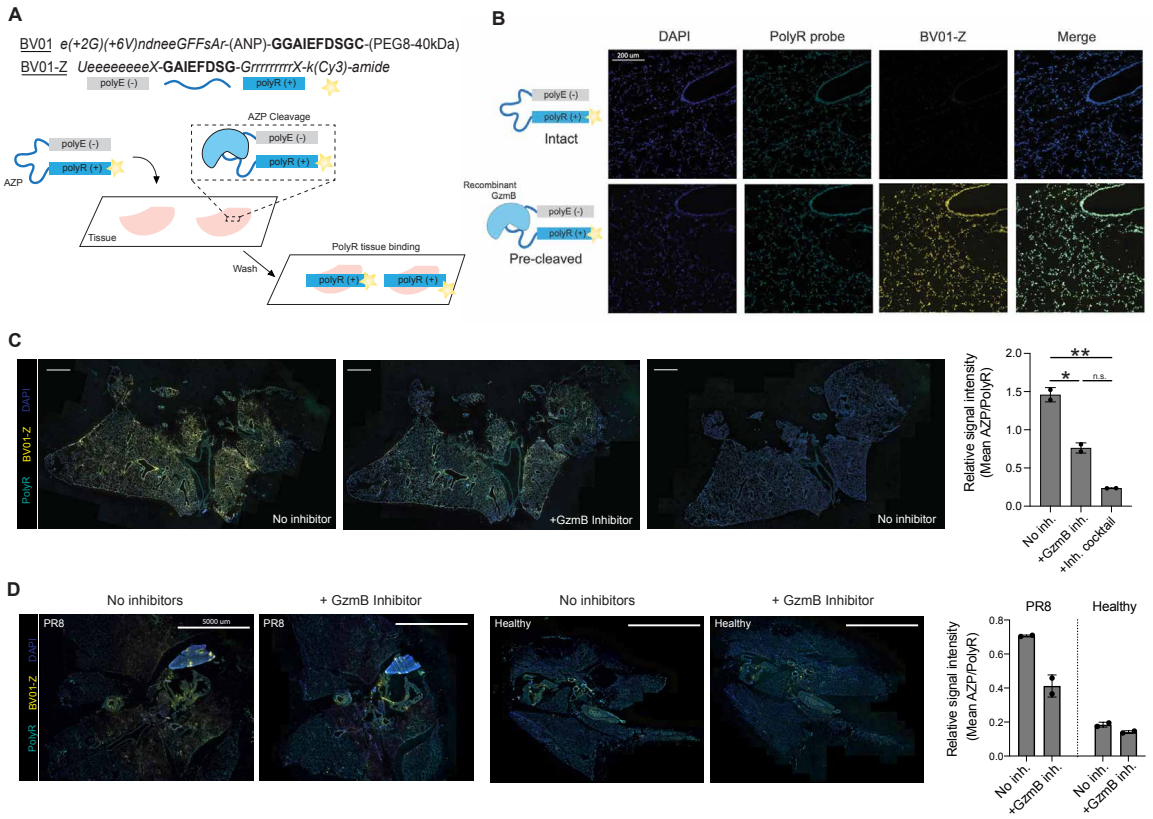


Figure 2-8: Granzyme B activity contributes to BV01 signal (A) The original BV01 substrate was incorporated into the activatable zymography probe (AZP) BV01-Z, consisting of the substrate sequence linking a fluorescently labeled polyR domain and a polyE domain. The AZP is applied to fresh frozen tissue and cleaved by active tissue-resident enzymes, after which the polyR domain electrostatically binds to the tissue. (B) Staining of fresh frozen healthy lung tissue with intact probe and BV01-Z that was incubated with GzmB before being applied to tissue. Staining shows free polyR (teal) and the cleaved polyR domain (yellow). (C) The GzmB responsive AZP BV01-Z (yellow) was applied to PR8 tissue with and without a GzmB specific inhibitor. Sections were co-stained with free polyR binding control (teal) and counterstained with DAPI (blue). Staining shows one full section on the slide with white squares showing the location of the zoomed in images. Scale bars for full sections = 1000 μm, scale bars for zoomed regions = 50 μm. Quantification of relative BV01-Z intensity in sections stained with or without GzmB inhibitor (n=2 consecutive sections, >60,000 detected cells per section, mean +/- s.d.; two-tailed unpaired t-test with Welch's correction, *p=0.0180). (D) Staining of fresh frozen lung from PR8 infected mice and healthy tissue with no inhibition or in the presence of a GzmB specific inhibitor. Each image is representative of consecutive sections. The ratio of the AZP signal (yellow) to free polyR (teal) is quantified (n=2 per condition).

there for up to 24 hours. [56] Therefore, the reporter enrichment that we observed *in vivo* is indicative of protease activity in the lungs, providing a readout of inflammation directly from the disease site.

Second, we created a nanosensor panel that can detect the differential activity of these proteases during pneumonia and release renally-cleared reporters to generate disease-specific signatures in urine. These ABNs have a plausible path to clinical approval and translation. The safety of the base formulation (i.e. a PEG scaffold with attached peptide linkers and conjugated mass barcodes) has already been established in both small and large animal models, as well as in healthy human subjects after intravenous administration in a phase I clinical trial. [82] Furthermore, our lab is developing aerosolized formulations of ABNs that are suitable for pulmonary delivery in patients. Finally, we have previously established lateral flow assays that can detect urinary ABN reporters in paper diagnostic formats. [83,84] These advances in safety, administration, and detection establish a path to develop our activity-based diagnostic platform for clinical use.

Importantly, in this work we have shown that a single panel of ABNs provides a readout of three different disease states. This is similar to techniques such as bacterial culture and multiplex PCR, but our ABNs have the added advantage of generating a readout within 2 hours of sensor administration, which is much faster than the hours to days required for other tests. Furthermore, the COVID-19 pandemic has highlighted some important limitations of PCR, such as the 24-48 hour turnaround time for results, the high cost per test, debates about the threshold for a positive result, and uncomfortable or invasive sample collection methods (e.g., nasopharyngeal swabs, sputum collection, bronchoalveolar lavage). It is also feasible that we could apply this same 20-plex panel to create signatures for other respiratory diseases, as we have shown that even within pneumonia the ABN signatures differ based on etiology. This is also supported by the fact that one ABN, BV07, was derived from our previous work in lung cancer. [56] Whereas BV07 signal was significantly decreased in lung cancer compared to healthy controls, it is increased in viral pneumonia.

Finally, we have identified a protease activity sensor, BV01, that contributes to etiology

stratification by detecting the activity of granzyme B (GzmB), which has increased expression in the mouse model of viral pneumonia than in bacterial pneumonia. This is further supported by the greater influx of GzmB-producing immune cells into the lungs of mice infected with PR8 versus during infection with *S. pneumoniae*. This shows that there are distinct differences in the immune response to viral versus bacterial pneumonia in terms of cell recruitment and subsequent protease expression, and that ABNs can be used to query these differences. In addition, creating AZP versions of other ABNs could yield a new way of detecting pneumonia *ex vivo*.

In summary, we have used our urinary reporter signatures to create diagnostic classifiers that can differentiate between bacterial and viral pneumonia and healthy controls with high accuracy. By focusing on the proteases implicated in host immunity, we have demonstrated that pneumonia diagnostics can leverage the body's innate response to pathogens to create noninvasive readouts of infection that are sensitive and specific. We believe that our activity-based nanosensor panel represents a new method of diagnosing pneumonia that could help disrupt the diagnostic paradigm.

2.4 Materials and Methods

2.4.1 Systematic search for gene expression datasets

We performed a systematic search in NIH Gene Expression Omnibus (GEO) and European Bioinformatics Institute (EBI) ArrayExpress for public human microarray genome-wide expression studies of TB or other diseases. [85, 86] Datasets were excluded if they (i) were nonclinical, (ii) were profiled using tissues other than WB or PBMCs, (iii) did not have at least 3 healthy samples, or (iv) did not provide information to identify whether a patient had bacterial or viral infection.

All microarray data were renormalized from raw data (when available) using standardized methods. Affymetrix arrays were renormalized using GC robust multiarray average (gcRMA) (on arrays with mismatch probes) or RMA. Illumina, Agilent, GE, and other

commercial arrays were renormalized via normal-exponential background correction followed by quantile normalization. Custom arrays were not renormalized. Data were log2-transformed, and a fixed-effect model was used to summarize probes to genes within each study. Within each study, cohorts assayed with different microarray types were treated as independent.

2.4.2 COCONUT conormalization

We conormalized data using COCONUT or Combat CONormalization Using conTrols. [59] COCONUT allows for conormalization of expression data without changing the distribution of genes between studies and without any bias towards sample diagnosis. It applies a modified version of the ComBat empirical Bayes normalization method that only assumes an equal distribution between control samples. [87] Briefly, the healthy controls from each cohort undergo ComBat conormalization without covariates, and the ComBat estimated parameters are acquired for each dataset's healthy samples. These parameters are then applied to the diseased samples in each dataset, which causes all samples to assume the same background distribution while still retaining the relative distance between healthy and diseased samples in each dataset. We have previously shown that when COCONUT conormalization is applied, housekeeping genes remain invariant across both diseases and cohorts, and each gene still retained the same distribution between diseases and controls within each data set.

2.4.3 Derivation of the 40-protease signature with MANATEE

MANATEE or Multicohort ANalysis with AggregaTed gEne Expression is a multicohort analysis framework that is used to integrate gene expression datasets, perform differential expression analyses to filter out top genes, apply machine learning methods to arrive at a concise diagnostic signature, and finally to validate the discovered signature in independent data (Fig. 2a). In this analysis, any genes that did not code for proteases were removed from all datasets. Next, relevant datasets were identified through a systemic search of public gene

expression data repositories. Some of these datasets were chosen for training the signature, and the rest were set aside as future independent validation datasets. Samples from the training datasets were then randomly split, with 70% of the samples assigned to Discovery and the other 30% assigned to Hold-out Validation. The Discovery and Hold-out Validation cohorts were each batch corrected with COCONUT conormalization.

Next, differential expression statistics were calculated in Discovery. Here, we computed four measures of differential expression between cases and controls are calculated for each protease: (1) the SAM score (from the Significance Analysis of Microarrays or SAM method) [88], (2) the corresponding SAM local FDR, (3) the Benjamini-Hochberg FDR corrected P value (from running a t-test [89]) and (4) the effect size (ES). The effect size is estimated as Hedges' adjusted g, which accounts for small sample bias. [90–93] We also performed a leave-one-study-out (LOSO) analysis, wherein each study that accounted for at least 5% of the training samples was iteratively removed from the training set, and the differential expression statistics were re-calculated for each version of the training set with one study left out. Thus, in order for a protease to be selected, it must not only exceed the given thresholds in the statistics calculated for the full training set, but it must also exceed those thresholds for each version of the training set with one study removed. This prevents any single study from exerting too strong of an effect on the selection of proteases. [94, 95]

Once the differential statistics were calculated, a set of "top" differentially expressed proteases was chosen by filtering out proteases that had an FDR of less than 0.01 and an absolute effect size of greater than 0.6. This resulted in a 40-protease signature. The signature was first tested in Hold-out Validation to assess whether the signature's performance remained robust when tested in new data. Finally, the signature was tested in Independent Validation to measure its performance in completely independent data.

2.4.4 Enrichment analysis with ConsensusPathDB

The bacterial and viral gene signatures were input into ConsensusPathDB for over representation analysis. For the pathway analysis there was a minimum overlap of 3 candidates

and a p-value cutoff less than 0.01. The entity graph visualization was performed using the database and edges with no shared candidates between nodes were filtered out.

2.4.5 Recombinant substrate screens with fluorescent substrates

Quenched fluorogenic probes were synthesized by CPC Scientific (sequences in table S3). Each probe was diluted first in dimethylformamide (DMF), subsequently in PBS, and plated into a 384-well plate. The plates were sealed and stored at -20°C until needed. To perform the cleavage assay, recombinant proteases were activated as necessary and diluted in their respective assay buffers with 0.1% BSA. The recombinant proteases were then added to each substrate containing well for a final reaction volume of 50 µL (20 µM substrate and 20 nM recombinant protease per well). Control wells, which contained no protease, were run on the same plate. Each protease-substrate pair and relevant blank control was plated in duplicate. Cleavage over time was quantified by fluorescence as measured by a fluorimeter (Tecan Infinite M200 Pro). Fold change was calculated as the fluorescent signal at 10 minutes divided by the original fluorescence at the start of the read. All enzyme sources and buffers can be found in Table 2.5.

2.4.6 Mouse pneumonia models

All animal studies were approved by the MIT IUCAC (protocol 0619-032-44) and were conducted in compliance with institutional and national policies. 7 to 9 week-old female mice (BALB/c, Taconic) were dosed with either *S. pneumoniae* (NCTC 7466), *K. pneumoniae* (ATCC 43816), *H. influenzae* (ATCC 33391), Pneumonia virus of mice (ATCC VR-1819), or influenza (Influenza A/PR/8/34 (H1N1), Charles River). The infectious dose for each pathogen was selected based on physical signs of infection in the mice and plated colony counts (for bacteria). To administer the pathogens, mice were first anesthetized by isoflurane inhalation (Zoetis). While under anesthesia, pathogens were passively inhaled via either intratracheal instillation (IT, for *S. pneumoniae*, *K. pneumoniae*, and *H. influen-*

zae) or intranasally (for PVM and Influenza A). A volume of 50 µL was administered for all pathogens except Influenza A, which was administered at 30 µL. Age- and gender-matched control mice in each experiment received either 50 µL of sterile-filtered PBS IT for the bacterial cohorts or IN for the viral cohorts.

2.4.7 Pathogen preparation

To prepare the bacteria, all bacteria were first cultured overnight (37°C, shaking at 250 rpm for 14-20 hours) and subsequently grown in secondary culture with 1:100 to 1:200 dilutions to an OD600 of 0.5-0.7, corresponding to a phase of exponential growth. *K. pneumoniae* was cultured in LB broth (Invitrogen). *S. pneumoniae* was plated overnight on blood-agar plates with neomycin (Hardy Diagnostics), and subsequently cultured in liquid brain-heart infusion (BHI; BD) media. *H. influenzae* was cultured in supplemented BHI (BHI with NAD and histidine-hemin). They were then pelleted, washed three times with sterile-filtered PBS and diluted to the appropriate concentration for administration. To prepare the viruses for infection, all viruses were diluted directly into sterile-filtered PBS from aliquoted stocks and kept on ice until administration.

2.4.8 qRT-PCR for viral loads and GzmB

Lungs were dissected from infected and healthy mice, rinsed in PBS and stored in RNAlater (Sigma Aldrich) at -80°C until use. RNA extraction was performed using the RNeasy Mini kit (Qiagen). On-column DNase digestion was performed using the RNase-Free DNase Set (Qiagen). RNA concentration was measured on a Nanodrop at A260. cDNA was prepared with the RevertAid First Strand cDNA synthesis kit (Thermo Fisher). qRT-PCR was performed using Ssofast EvaGreen Supermix (Bio-Rad). For viral load quantification, custom oligo primers for PR8 and PVM were ordered from IDT (PR8: PA gene, Forward: 5' GCG GTC CAA ATT CCT GCT GA 3', Reverse: 5' CAT TGG GTT CCT TCC ATC CAA AG 3'; PVM: SH gene, Forward: 5' GCC GTC ATC AAC ACAG TGT GT 3', Reverse: 5' GCC

TGA TGT GGC AGT GCT T 3'). Viral loads were estimated by running a standard curve with custom gBlocks from IDT for PR8 (PA gene) and PVM (SH gene). Relative Granzyme B expression was measured using custom primers (IDT, Forward: 5' TCT CTG ACT CCA CGT CTC TTA C 3', Reverse: 5' CTG GGT CTT CTC CTG TTC TTT G 3'). GAPDH expression was measured for normalization (ReadyMade primers, IDT).

2.4.9 In vivo activity-based nanosensor studies

Nanosensors were synthesized by CPC Scientific. ABNs were dosed in mannitol buffer (0.28 M mannitol, 5 mM sodium phosphate monobasic, 15 mM sodium phosphate dibasic, pH 7.0-7.5) and deposited into the lungs by intratracheal instillation (50 µL total volume, 20 µM per ABN). Immediately after dosing, all mice were given a subcutaneous injection of PBS (400 µL) to promote adequate urine volumes for subsequent analysis. For the viral pneumonia models, mice were administered the ABN cocktail 6 days post infection (p.i.). For the bacterial pneumonia models, mice were administered ABNs 16 hours p.i. For all mice, after receiving the ABNs mice were returned to their home cage for one hour with full access to food and water. After this hour their bladder was manually voided, and they were transferred into a urine collection chamber. At the end of the second hour, the bladder was manually voided and the urine was collected, along with any urine that was produced in the collection chamber. The urine samples were then sent to Syneos Health for LC-MS/MS analysis. Reporter quantification by LC-MS/MS was performed as previously described. [56]

2.4.10 Tissue dissection from mice and slide preparation

Female BALB/c mice were infected with influenza A (PR8) or *S. pneumoniae* (SP) as described above. PR8 and SP mice were euthanized at 6 days and 16 hours after infection initiation, respectively. The lungs were removed from the infected mice or healthy controls, and put into a 6-well plate filled with PBS while the lobes were separated. The individual

lobes were then immediately embedded in optimal-cutting-temperature (OCT) compound (Sakura), frozen in isopentane chilled with dry ice, and stored at -80ºC until sectioning. Cryosectioning was performed at the Koch Institute Histology Core. The resulting slides were then stored at -80ºC until use for immunofluorescent staining or AZP experiments.

2.4.11 Immunofluorescent staining for immune cell markers and GzmB

Fresh frozen slides were prepared as described above. To prepare the slides for staining, they were air-dried for 20 minutes, fixed in ice-cold acetone for 10 minutes, air-dried for 20 minutes and washed in sterile PBS (3x5 minutes). After the final wash, the tissue sections on each slide were outlined with an ImmeEdge Pen (Vector Laboratories Inc, Bulingame, CA) and blocked with 1% BSA in PBS for 30-45 minutes. The blocking buffer was then aspirated and replaced with the relevant primary antibodies in PBS (Granzyme B, Abcam 25598, 2.93 μg/mL; RB6-8C5, Abcam 25377, 5 μg/mL; Mouse NKp46/NCR1 Antibody, R&D AF2225, 4 μg/mL; CD8 (53-6.7), Novus NBP1-49045, 5 μg/mL), and incubated at room temperature for 1.5 hours. Slides were washed with PBS (3x5 minutes) and incubated with appropriate secondary antibodies and Hoechst (1:2000 dilution) for 30 minutes at room temperature. They were then washed (3x5 minutes) and mounted (ProLong Diamond Antifade Mountant, Invitrogen).

2.4.12 In situ zymography with AZPs

For experiments involving on-slide AZP activation, slides were dried and fixed as previously described. Slides with Granzyme B inhibitor (Z-AAD-CH2Cl, Abcam ab142034) were blocked with 1% BSA and 100 μM inhibitor in PBS, and those without were blocked with 1% BSA and an equivalent volume of DMSO to the GzmB inhibitor. After blocking, slides were incubated with BV01-Z (1 μM), Cy7-polyR (1 nM), and either GzmB inhibitor (100 μM) or equivalent DMSO volume diluted in buffer (50 mM Tris, pH 7.5) for 2 hours at 37ºC. Slides were washed with PBS (3x5 minutes) and incubated with Hoechst (1:2000 dilution)

for 10 minutes at room temperature. Slides were washed again (3x5 minutes) and mounted as previously described. For pre-cleavage experiments, recombinant human Granzyme B (R&D 2096-SE, 100 μg/mL) was activated with recombinant mouse Cathepsin C (R&D 2336-CY-010, 10 μg/mL) in activation buffer (50 mM MES, 50 mM NaCl, pH 5.5) for 4 hours at 37°C. The activated rhGzmB was diluted to 100 nM and incubated with BV01-Z (10 μM) and DNTB (100 μM) in assay buffer (50 mM Tris, pH 7.5) for 4 hours at 37°C. Meanwhile, fresh frozen slides with healthy lung tissue were prepared and blocked with BSA as described. After blocking, the slides were either incubated at 4°C for 1 hour with the pre-cleaved BV01-Z mixture or intact BV01-Z, Cy7-polyR and DNTB diluted in assay buffer. Slides were then washed (3x5 minutes), stained with Hoechst, washed, and mounted as previously described.

2.4.13 Quantification of immunofluorescent staining and AZP signal

All slides were imaged on a Pannoramic 250 Flash III whole slide scanner (3DHistech). Whole slide images were imported into QuPath (0.2.3) for quantification. Individual cells were detected using the cell detection feature on the DAPI channel. Intensity thresholds were manually determined based on mean and maximum intensity distributions for each channel, in order to classify cells as being either positive or negative for any given marker (CD8, NK, GzmB, Ly6G). Using scripts, each cell on the slide was annotated based on whether it met the threshold, and the percentage of positive cells for each marker was calculated using Excel. For AZP quantification, the mean intensity of each cell in the AZP and polyR channels was calculated and the ratio of those values was interpreted as the relative AZP signal. All further statistical measurements were performed in GraphPad 9.0 (Prism).

2.4.14 Statistical analysis

Principal component analysis, reporter enrichment, and the support vector machine algorithm training were performed in the Python with the scikit-learn environment (https://scikit-

learn.org). Hierarchical clustering was performed in R (https://www.r-project.org/). All other analyses were performed in GraphPad 9.0 (Prism).

2.5 Author contributions

Sections of this work were derived from Anahtar et. al (in review). As noted in that paper:

M.A. conceived the study design with feedback from Leslie W. Chan (L.W.C.) and Sangeeta N. Bhatia (S.N.B.), and performed *in vitro* experiments. Aditya Rao (A.R.), M.A., and Purvesh Khatri (P.K.) created the transcriptomic signature. M.A., Henry Ko (H.K.) and L.W.C. optimized the CAP models. M.A., L.W.C., and H.K. performed *in vivo* experiments. M.A. and A.P.S. analyzed the data. M.A. wrote the paper with contributions from A.P.S. and A.R., and feedback from all authors.

2.6 Acknowledgements

We thank H. Fleming for her thoughtful feedback on the manuscript, J. Kirkpatrick, T. Samad, L. Gehrke and D. Hung for their technical advice, A. Mancino (Syneos Health) for performing mass spectrometry, and the Koch Institute Swanson Biotechnology Center Histology Core for slide processing. This work was supported in part by the Global Health Innovation Partnership Sentinel Award from the Bill and Melinda Gates Foundation and Janssen Research & Development. M.A. acknowledges support from the National Science Foundation Graduate Research Fellowship Program (GRFP). L.W.C. acknowledges support from the National Institutes of Health Pathway to Independence Award (K99 EB28311). A.P. Soleimany acknowledges support from the NIH Molecular Biophysics Training Grant (S) and the National Science Foundation GRFP. S.N.B. is a Howard Hughes Institute Investigator.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 3

Designing diagnostic classifiers for pneumonia

3.1 Introduction

Artificial intelligence (AI) algorithms have become a quintessential component of the diagnostic paradigm. Researchers and clinicians around the world have been incorporating AI into their work in the hopes that computers can prevail where humans fail. Machine learning (ML) is an application of AI whereby an algorithm can learn to produce a behavior that has not been explicitly programmed, thus allowing the machine to gain human-like intelligence. It is particularly useful in making predictions based on large datasets, therefore we sought to use ML to translate the urinary signatures of our 20-plex ABN panel into diagnoses of disease. To do this, we used a ML model called a Support Vector Machine (SVM), which can assign unknowns into different categories. In our case, these categories are pneumonia versus healthy, and bacterial versus viral pneumonia.

To do this, labeled objects from each group are fed into the model to train the algorithm. The algorithm then essentially plots the objects in n-dimensional space, where n is the number of features for each object, and creates a mathematical function that optimizes separation between the two categories of objects. This function is termed the "separating

hyperplane", and is the crux of SVM's usefulness. The model is then fed new datapoints, comprising the "test" dataset, which are then classified as members of group 1 or group 2 based on where each point falls relative to the separating hyperplane. This process can then be iterated upon to distinguish between multiple categories, which is termed multiclass classification.

Previously, we had shown clear separation between pneumonia and healthy mice, as well as bacterial versus viral etiology, via PCA. This separation in two-dimensional space indicated that there were clear differences in the urinary reporter concentrations among mice with varied disease states. To translate the differences among all 20 nanosensors into something that could be diagnostically useful, we leveraged SVM to create classifiers that can simultaneously detect CAP and distinguish etiology with high specificity and sensitivity. These classifiers were also able to distinguish between pneumonia and healthy mice when applied to pathogens that were not used for training, demonstrating that they are generalizable. We also showed that high accuracy can be maintained when limiting the panel to just five ABNs, which puts less technical readout modalities such as lateral flow assays within reach. Thus, with our ABN panel, we have created a proof of concept for a non-invasive urinary diagnostic for CAP with broad clinical utility.

3.2 Results

3.2.1 The ABN panel can classify pneumonia and determine etiology

Ultimately, our goal is to use our protease activity sensor panel to noninvasively diagnose pneumonia and determine etiology without any prior knowledge of a subject's infectious state. Our enrichment analysis demonstrated that ABNs were differentially cleaved based on disease state (Fig. 2-4), and our AZP results gave us confidence that the reporters present in the urine were reflective of protease activity that was driven by disease pathogenesis (Fig. 2-8). Confident that the urinary reporters were reflective of disease, we sought to leverage their relative concentrations to create a diagnostic tool. For this, we leveraged ma-

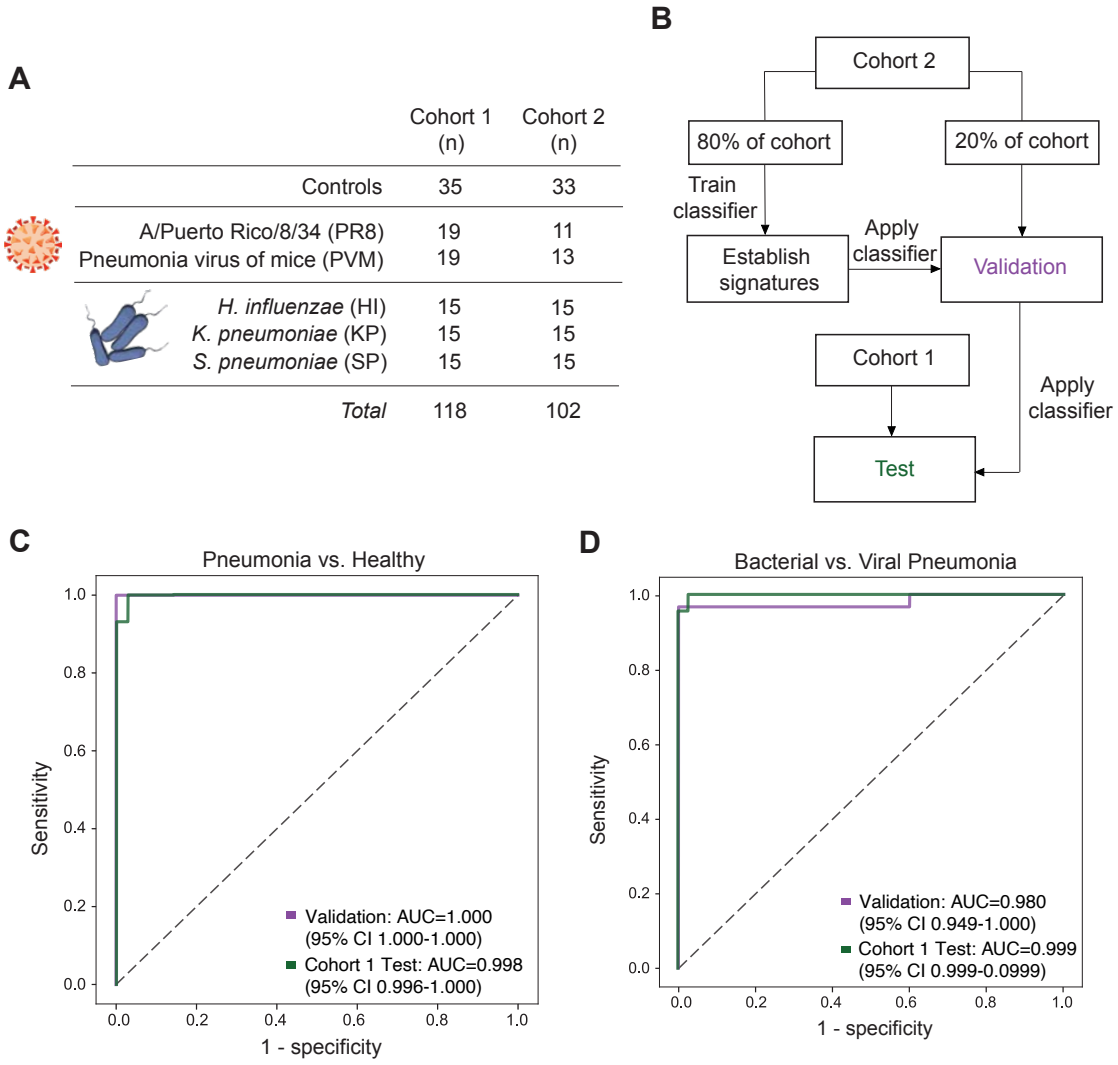


Figure 3-1: The panel can diagnose pneumonia and distinguish pneumonia etiology with high accuracy. (A) Mice from two independent cohorts were infected with various pneumonia-causing pathogens or given mock PBS for healthy controls. (B) Flowchart of the training and validation method used to create and test support vector machine (SVM) classifiers for pneumonia. Cohort 2 was split into two groups, one to train the classifier and another to validate its performance. The classifier was then applied to a separate group of infected and healthy mice, Cohort 1. The classification performance of the classifier on this independent cohort is labeled as the test condition. The validation and test performance of binary classifiers trained using this framework is represented with ROC curves. Distinguishing power of classifiers trained on a multiclass algorithm were visualized with a confusion matrix. (C, D) Performance of binary classifiers to differentiate mice infected with pneumonia from healthy controls (C) and bacterial from viral pneumonia (D). All classifiers are averages over 10 independent train-test trials. Train, validation, and test n can be found in the text.

chine learning to build a classifier that could be prospectively applied to enable pneumonia diagnosis. To do so, we infected a completely new set of mice (Cohort 2, n=102). These mice infected with the same pathogens as the original cohort (Cohort 1, n=118) but all aspects of the infection, ABN administration, and urine collection processes were performed independently. We then split Cohort 2 into one group (n=81) that was used to train a support vector machine (SVM) classifier and another (n=21) to validate the classifier's performance (Fig. 3-1b). The classifier was able to perfectly distinguish between mice that had pneumonia and healthy controls within this validation group (AUC 1.0, Fig. 3-1c). We then took this classifier and applied it to the urine reporter concentrations from Cohort 1 to evaluate its diagnostic potential. The classifier achieved near perfect classification (AUC = 0.998, Fig. 3-1c), thus demonstrating that the ABN panel can be used to train a classifier that can diagnose pneumonia.

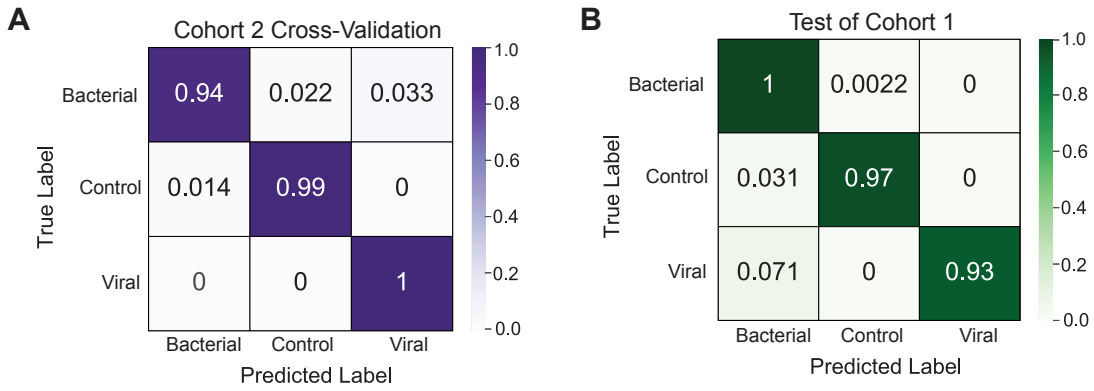


Figure 3-2: Confusion matrices visualizing performance of a multiclass SVM algorithm to distinguish between all 3 states of interest. Each value represents the frequency at which each true label was classified with the predicted label (e.g., the top left box represents the mice with bacterial pneumonia, the true label, that were classified as having bacterial pneumonia, the predicted label). The diagonal represents the true positive classifications. Train, validation, and test n can be found in the text. (A) Cross-validation of the diagnostic classifier derived from the Cohort 2 training group was performed on a subset of a samples and (B) tested on cohort 1. Values in the matrix represent the percentage of mice that were assigned each label in relation to the true identity of the disease state for each mouse. All classifiers are averages over 10 independent train-test trials.

Using the same framework, we then used the reporter concentrations from the infected population of Cohort 2 (n=69) to train and validate a new binary classifier to distinguish

viral and bacterial pneumonia. Again, we observed very high performance in the validation cohort (AUC = 0.980, Fig. 3-1d). To test the bacterial-viral classifier, we applied it to the infected mice from Cohort 1 (n=83) and saw near perfect classification of the infected mice based on pneumonia etiology (AUC = 0.999, Fig. 3-1d). After establishing that binary classifiers could be used to separately identify pneumonia and distinguish its cause, we then sought to put these capabilities together. To do this, we created a multiclass SVM algorithm that was trained and validated on Cohort 2 and then tested on Cohort 1. As expected, the urinary signatures could differentiate between these three states with near perfect accuracy (Cohort 2 validation Fig. 3-2a, Cohort 1 test Fig. 3-2b). These results demonstrate that our ABN platform can be used to not only identify pneumonia but also stratify etiology.

After establishing that classifiers could be trained to differentiate between viral and bacterial pneumonia, we next sought to evaluate whether the urine signatures were unique enough to enable pathogen specific identification. PCA analysis of Cohorts 1 and 2 showed that the pathogens formed fairly distinct clusters (Fig. 3-3a). By training a multiclass classifier on Cohort 2 and testing it on Cohort 1, we demonstrated that a classifier could be created to correctly identify the pneumonia-causing pathogen, or lack thereof in the case of the healthy controls, in each mouse with high accuracy.(Fig. 3-3b,c) In cross-validation, the classifier struggled slightly with distinguishing between the two viral models. This was to be expected as the PCA of Cohort 2 reveals a large amount of overlap between PVM and PR8, suggesting that the two viruses produce similar urinary signatures. Importantly, this error would not affect classification when focusing on distinguishing bacterial versus viral etiology, as ultimately this result shows that the viral ABN signatures are similar, yet distinct from the bacterial signatures. Applying this classifier to the test set also yields highly accurate identification for all pathogens except HI, which can be mislabeled as healthy controls. Given these errors, it is clear that while the ABN panel does enable pathogen identification, further optimization is required to reach a level of accuracy that would enable microbe-directed treatment.

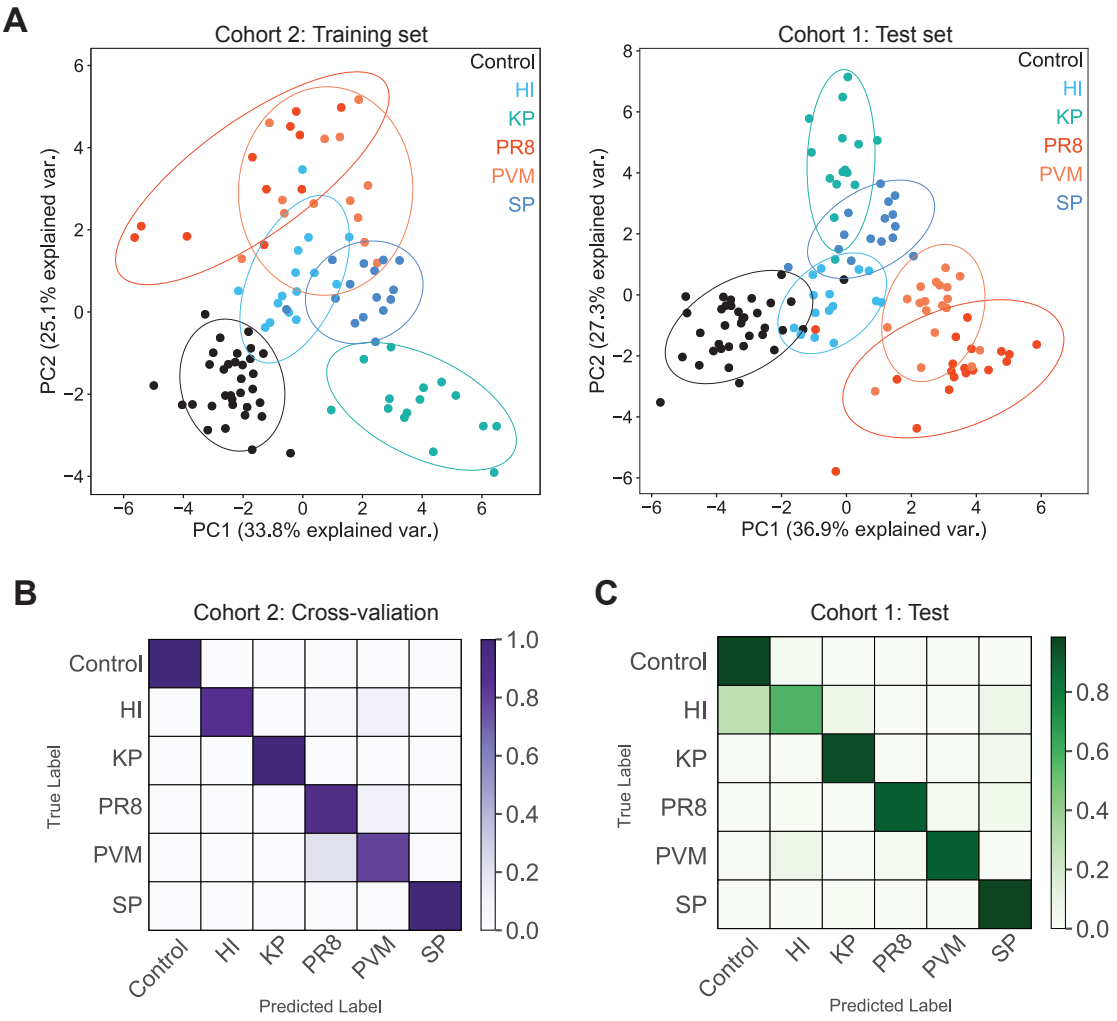


Figure 3-3: A multiclass SVM algorithm can be trained to identify each pneumonia-causing pathogen. (A) Principal component analysis (PCA) was performed on mean normalized urinary ABN signals from healthy controls (black) and mice with pneumonia (colored symbols) in Cohorts 1 (right) and 2 (left). (B,C) Confusion matrices showing the accuracy of a support vector machine (SVM) classifier in distinguish each pathogen from one another. All classifiers are averages over 10 independent train-test trials.

3.2.2 The ABN panel can be reduced without sacrificing accuracy

Multiplexing our nanoparticles enables the creation of specific urinary signatures that can reflect subtle differences between disease states. However, point-of-care diagnostics typically do not have strong multiplexing capabilities, as tests must be simplified for universally accessible sampling, processing, and analysis. Therefore, while our 20-plex panel was able to distinguish between pneumonia etiology with high sensitivity and specificity, to demonstrate that this panel has potential to be used as a point-of-care diagnostic, we sought to determine the minimal subset of ABNs that could still achieve high classification. To do this, we re-performed the analysis while iterating through various combinations of the reporters to determine which minimal set could still achieve differentiation. Creating a classifier using the top two differentially expressed reporters between bacterial and viral pneumonia (BV03 and BV19) yielded fairly robust classification using a binary classifier (AUC=0.951, Fig. 3-4a). This is promising in terms of creating a simple test that can guide antibiotic stewardship if a patient is already known to have pneumonia but the underlying cause needs to be identified, as measuring the presence of two analytes can be done using fairly straightforward lateral flow assays and other traditional point-of-care tests. However, testing using a multiclass classifier showed that this reporter pair could not classify etiology in the wider context of comparing to healthy controls, making the pair unsuitable for use in a more classic clinical setting. Based on the differential enrichment analysis performed in chapter 2 (Fig. 2-4c,e), we gradually incorporated other reporters into the set and determined that a set of five ABNs (BV19, BV03, BV13, BV09, and BV07) could determine etiology with high accuracy using both binary and multiclass classifiers (Fig. 3-4d).

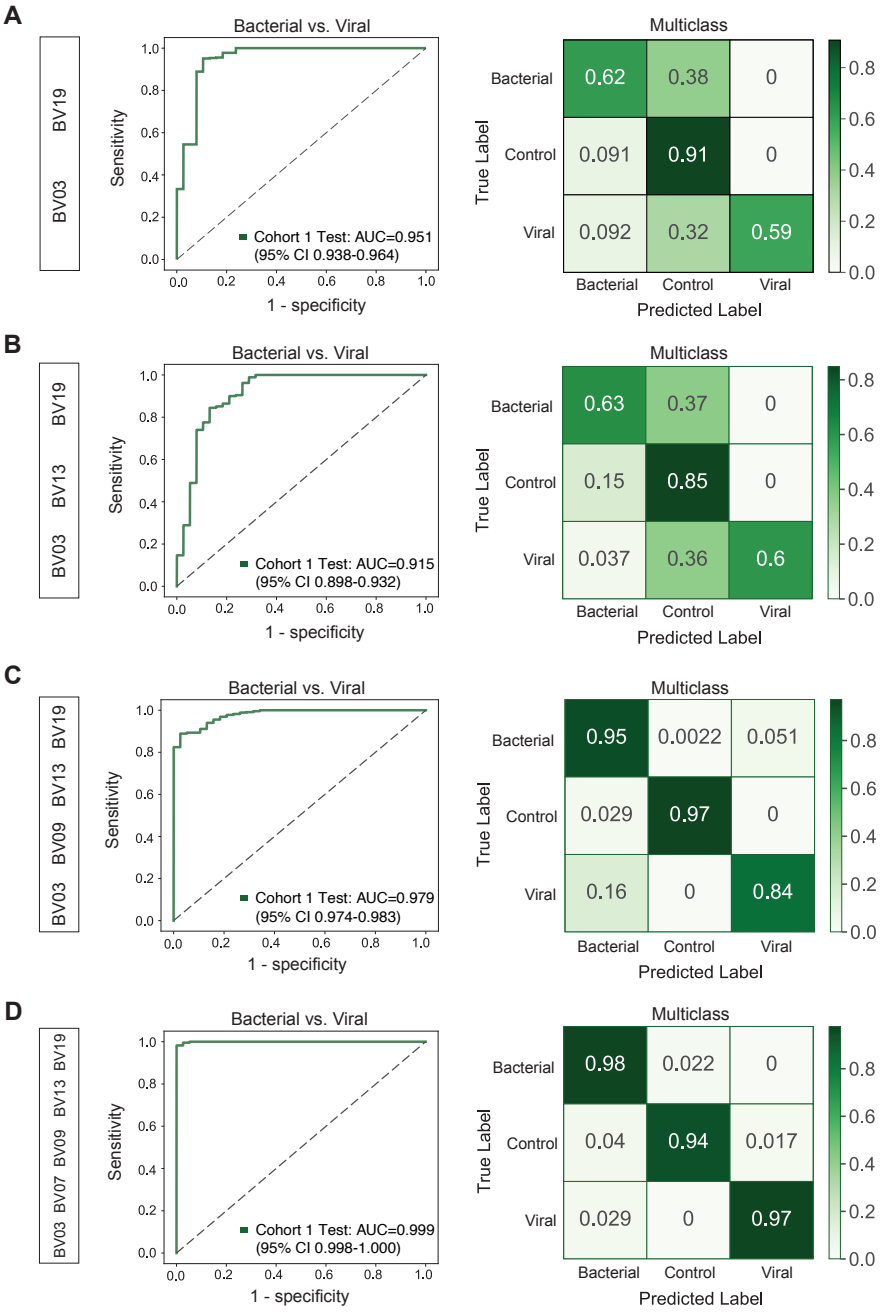


Figure 3-4: A subset of five ABNs can achieve high binary and multiclass classification of etiology. ROC curves and confusion matrices showing the performance of a support vector machine trained on urinary reporters from the mice in Cohort 2. Each pair of graphs shows the performance after training classifiers using the changing set of reporters listed on the left. All classifiers are averages over 10 independent train-test trials.

3.2.3 The urinary signatures are generalizable to causes of CAP beyond the training set

We next sought to evaluate whether the ABN panel produced urinary signatures that could be generalizable to pathogens that were not used to train the diagnostic classifier, as the five pathogens that were tested using the ABNs represent a small subset of the bacteria, viruses, and fungi that can cause CAP. Using our existing data, we tested this by systematically excluding pathogens from the training set, while testing the resulting classifier on all five pathogens (Fig. 3-5). This exclusion analysis revealed that pneumonia virus of mice (PVM) and *H. influenzae* (HI) were the most important pathogens to include in the training set in order to achieve high quality classification, as indicated by the fact that excluding them from the training set led to misclassification of viral pneumonia as bacterial, and bacterial pneumonia as healthy controls, respectively.

Based on that finding, we then trained a classifier using urine data from mice infected with HI, PVM, and healthy controls, and found that the resulting classifier could reliably predict the etiology of the remaining three pathogens (*S. pneumoniae*, *K. pneumoniae*, and influenza; Fig. 3-5g). This indicates that the urinary signatures generated by bacterial and viral pneumonia are somewhat generalizable, as a single representative pathogen from each group can be used to distinguish broad differences in protease activity between etiologies. Most promisingly, this holds even when the cohort used for training and testing is reversed (Fig. 3-5i), showing that this effect is not biased by cohort selection. We then sought to combine this minimal pathogen set with the minimal reporter set that we previously identified. Unsurprisingly, we found that limiting the ABN panel from the original 20-plex to the previous 5-plex strongly impacted the accuracy of the resulting classifier Fig. 3-5j). This finding validates our previous results, in the sense that it shows that multiplexing can improve performance by creating more nuanced signatures of disease.

Having established that the ABN panel and resulting diagnostic classifiers could be used to generally discern bacterial and viral pneumonia, and even identify specific disease causing pathogens, we then wanted to evaluate its performance on other CAP-causing pathogens.

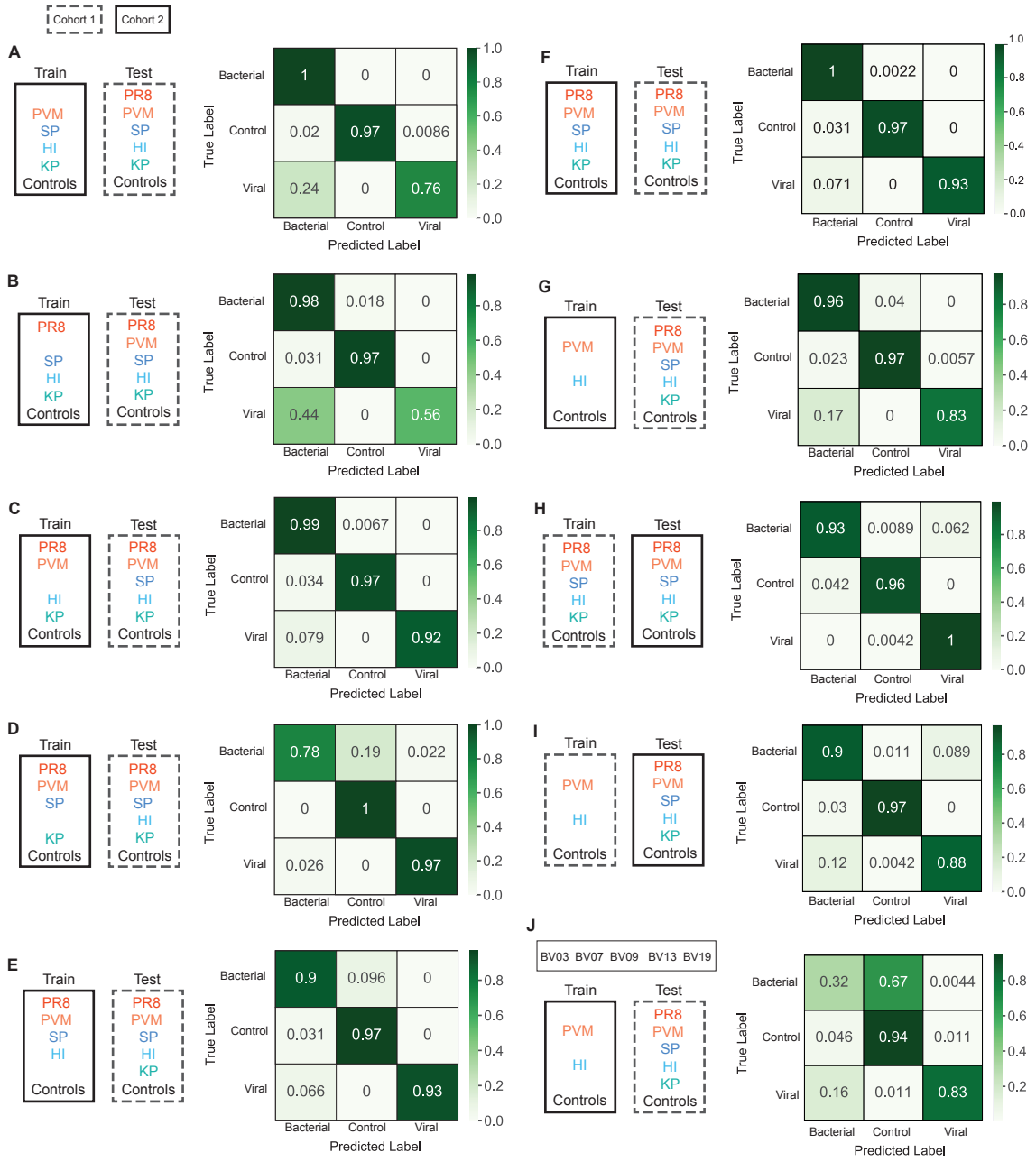


Figure 3-5: Classifier performance can be maintained after training on limited sets of disease states. Confusion matrices displaying the performance of a multiclass classifier to distinguish between bacterial and viral pneumonia, and healthy controls. The pathogens included in each training set were varied, as were the cohorts that the data was derived from. The resulting classifiers were then tested on all pathogens in an independent cohort. In the final panel (J), the classifier was trained on a limited set of disease states and a narrowed set of ABN reporters. All classifiers are averages over 10 independent train-test trials.

Staphylococcus aureus is another common cause of CAP and is a gram-positive bacteria that colonizes the upper airways in healthy humans, making it similar to *Streptococcus pneumoniae*. However, despite these similarities, studies have shown that pneumonia caused by *S. aureus* produces a distinct metabolic profile compared to *S. pneumoniae*, suggesting that though these two bacteria share many qualities, they influence the host response in unique ways. After optimizing the *S. aureus* pneumonia model (Fig. 3-6a), we administered the full ABN panel into mice with *S. aureus* (SA) pneumonia and parallel healthy controls (Cohort 3, n=10 controls, n=13 SA). PCA showed distinct separation between the two groups, similar to what had been previously observed with the original panel of bacteria (Fig. 3-6b). We then used the previously derived diagnostic classifier, tested it on Cohort 3, and observed that there was perfect classification of healthy versus pneumonia mice.(Fig. 3-6c) This indicated consistency among the urine signatures produced by the control mice, as this was a completely independent set of mice, and among the mice with pneumonia, as this new pathogen was not included in the training cohort. To evaluate how the *S. aureus* derived urinary signatures compared to the other pneumonia-causing pathogens, we compiled the data from cohorts 2 and 3 and re-performed PCA. Surprisingly, we observed that *S. aureus* overlapped significantly with the viral pneumonia models and distinctly from the other bacteria (Fig. 3-6d).

To determine whether this viral-like signature was unique to *S. aureus*, we then optimized and tested another mouse model of bacterial pneumonia, this time infecting with *Pseudomonas aeruginosa* (PA). This bacteria is a gram-positive pathogen, like *H. influenzae* and *K. pneumoniae*. For this experiment, we administered the ABN panel alongside a group of healthy CD-1 mice, in order to simultaneously assess whether the panel would behave differently in a mouse strain with a distinct genetic background (Cohort 4, n=20 healthy CD-1, n=18 PA). Once again, we observed distinct separation between the healthy CD-1 mice and the *P. aeruginosa* infected BALB/c (Fig. 3-6f), and demonstrated that the binary classifier derived from Cohort 2 could correctly diagnose both sets of mice as pneumonia versus healthy, despite the distinct characteristics of both groups (Fig. 3-6g). However, when we

compared Cohorts 2 and 4 using PCA we observed that unlike *S. aureus*, *P. aeruginosa* did not overlap with any other pathogen; rather it formed its own distinct group (Fig. 3-6h). In contrast, the CD-1 group overlapped with the BALB/c controls.

Based on the PCA results for both *S. aureus* and *P. aeruginosa*, we were unsurprised to find that using the multiclass classifier trained on Cohort 2, both pathogens were frequently misidentified as viral pneumonia (Fig. 3-6i,j). However, we hypothesized that by including Cohort 4 in the training set, we could improve classifier performance for *S. aureus*, as the classifier would be exposed to bacterial signatures with strong viral "components". Indeed, we observed this to be the case (Fig. 3-6k), indicating that the composition of the training set can greatly influence classification power.

3.3 Discussion

Artificial intelligence is commonly employed to detect abnormalities in chest x-ray and computed tomography (CT) images with a level of precision that is unmatched by the human eye. This has made it an increasingly popular area of research throughout the COVID-19 pandemic, and studies have shown that AI models can be trained to diagnose COVID-19, and differentiate it from other viral and bacterial pneumonias. While computer-aided diagnosis of chest images is undoubtedly useful, an estimated two-thirds of the world's population does not have access to diagnostic imaging, making the clinical utility of these algorithms moot. Activity-based nanosensors (ABNs) offer an alternative means of diagnosing pneumonia. In this work, we have shown that machine learning algorithms can be trained on urinary signatures generated by the 20-plex to create diagnostic classifiers that both diagnose pneumonia, distinguish etiology, and identify specific disease-causing pathogens. We have also identified a set of five ABNs from the original 20-plex that can characterize etiology with high accuracy, which primes the pneumonia ABNs for future development using lower-throughout readout modalities.

The COVID-19 pandemic has made it clear that determining the etiology of pneumo-

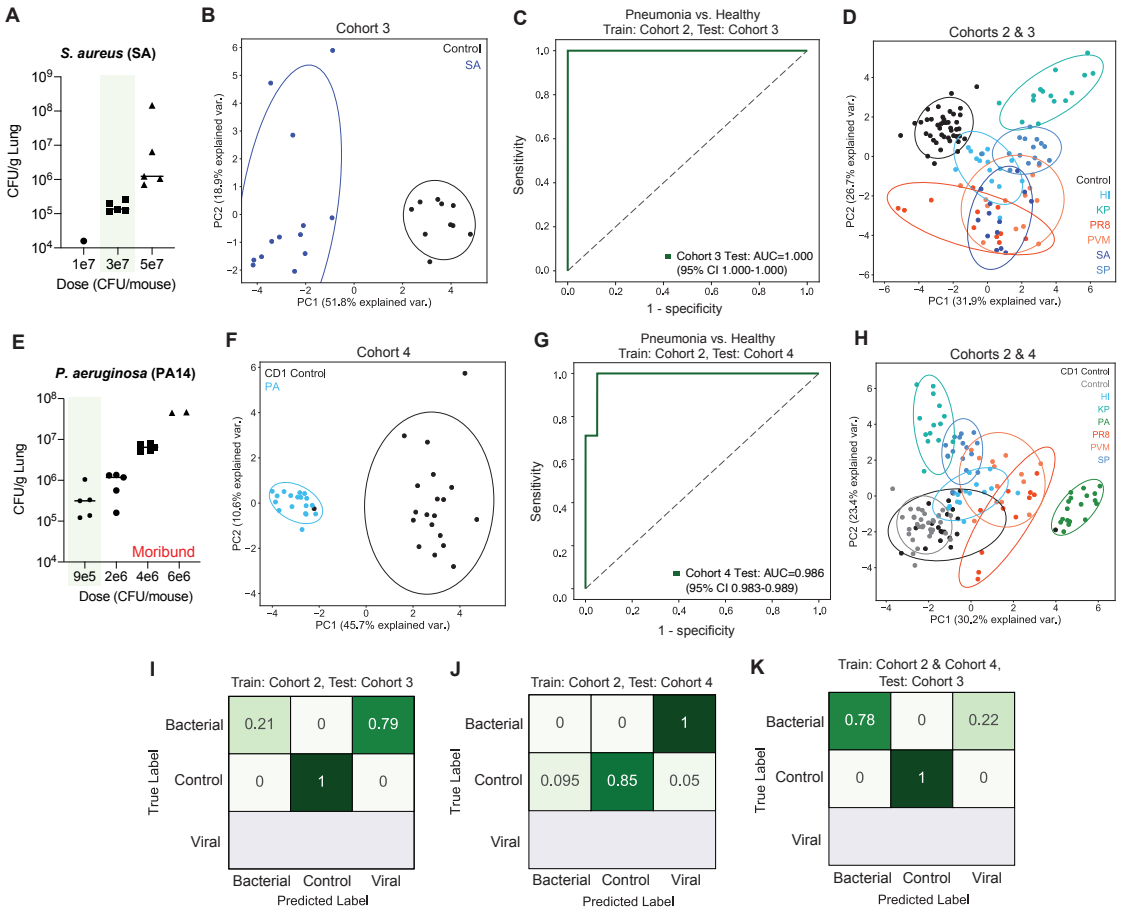


Figure 3-6: Diagnostic classifiers can be generalized to urinary signatures from pathogens outside of the training set. (A,E) Mouse models of *S. aureus* (SA) and *P. aeruginosa* (PA) were optimized. (B, F) PCA of independent mouse cohorts that were administered the 20-plex ABN panel. Cohort 3: BALB/c controls (n=10), SA (n=13); Cohort 4: CD-1 controls (n=20), PA (n=18). (C,G) ROC of performance for a classifier trained on Cohort 2 and tested on Cohorts 3 (C) and 4 (G). (D,H) PCA of Cohorts 3 and 4 pooled with Cohort 2. (I,J) Confusion matrices of performance by a three-way classifier trained on Cohort 2 and tested on Cohorts 3 (I) and 4 (J). (K) Performance of a classifier trained on Cohorts 2 and 4, and tested on Cohort 3. All classifiers are averages over 10 independent train-test trials.

nia is crucial for individual patient management and public health. Here we have used our urinary reporter signatures to create diagnostic classifiers that can differentiate between bacterial and viral pneumonia and healthy controls with high accuracy. By focusing on the proteases implicated in host immunity, we have demonstrated that pneumonia diagnostics can leverage the body's innate response to pathogens to create noninvasive readouts of infection that are sensitive and specific. We believe that our activity-based nanosensor panel represents a new method of diagnosing pneumonia that could help disrupt the diagnostic paradigm.

While the reliance mass-spectrometry analysis does not make ABNs a feasible diagnostic tool in middle- and low-income settings, the identification of a five reporter panel that can diagnose pneumonia puts point-of-care use within reach. Lateral flow assays (LFAs) are ideal diagnostic tools for point-of-care use in resource limited settings due to their low cost, quantification potential, and user friendliness. Our lab has previously shown that LFAs can be used to read out ABN activity *in vivo* [83] for up to four urinary reporters. However, there are multiplexing strategies for LFAs that can enable the detection of 10-30 target simultaneously. Therefore, the set of five ABNs that we identified is well within the design parameters for the future design of an LFA for diagnosing pneumonia.

An important biological limitation of this work is that a wide range of bacterial and viral pathogens cause CAP. Thus, to use the ABN panel as a point-of-care diagnostic, the classifier will need to be trained on urinary signatures derived from human subjects infected with pathogens beyond the seven included in this work. For example, emerging causes of CAP, such as SARS-CoV-2, should be included in classifier training. Promisingly, we have shown that the urinary signatures generated by a small subset of CAP causing pathogens can be used to train classifiers that are generalizable to novel pathogens. Furthermore, recent work has shown that respiratory viruses such as influenza, SARS-CoV-2, Ebola, and others have well conserved dysregulation of the host immune response [96], which supports that our diagnostic may be generalizable to novel and evolving viruses. In addition, the inclusion of data from pathogens that confounded the classifier when they were not included in training

(e.g. *P. aeruginosa*), should help boost classifier performance.

3.4 Materials and Methods

3.4.1 Mouse pneumonia models

All animal studies were approved by the MIT IUCAC (protocol 0619-032-44) and were conducted in compliance with institutional and national policies. 7 to 9 week-old female mice (BALB/c, Taconic) were dosed with either *S. pneumoniae* (NCTC 7466), *S. aureus* (502A), *K. pneumoniae* (ATCC 43816), *H. influenzae* (ATCC 33391), *P. aeruginosa* (PA01, PA14), pneumonia virus of mice (ATCC VR-1819), or influenza (Influenza A/PR/8/34 (H1N1), Charles River). The infectious dose for each pathogen was selected based on physical signs of infection in the mice and either plated colony counts (for bacteria) or viral titers. To administer the pathogens, mice were first anesthetized by isoflurane inhalation (Zoetis). While under anesthesia, pathogens were passively inhaled via either intratracheal instillation (IT, for *S. pneumoniae*, *S. aureus*, *K. pneumoniae*, and *H. influenzae* and *P. aeruginosa*) or intranasally (for PVM and Influenza A). A volume of 50 µL was administered for all pathogens except Influenza A, which was administered at 30 µL. Healthy age and gender-matched control mice in each experiment received either 50 µL of sterile filtered PBS intratracheally for the bacterial cohorts or 50 µL of sterile filtered PBS intranasally for the viral cohorts.

3.4.2 Pathogen preparation

To prepare the bacteria, all bacteria were first cultured overnight (37°C, shaking at 250 rpm for 14-20 hours) and subsequently grown in secondary culture with 1:100 to 1:200 dilutions to an OD600 of 0.5-0.7, corresponding to a phase of exponential growth. *K. pneumoniae*, *S. aureus*, and *P. aeruginosa* were cultured in LB broth (Invitrogen). *S. pneumoniae* was plated overnight on blood-agar plates with neomycin (Hardy Diagnostics), and subsequently cultured in liquid brain-heart infusion (BHI; BD) media. *H. influenzae* was cultured in sup-

plemented BHI (BHI with NAD and histidine-hemin) They were then pelleted, washed three times with sterile-filtered PBS and diluted to the appropriate concentration for administration. To prepare the viruses for infection, all viruses were diluted directly into sterile-filtered PBS from aliquoted stocks and kept on ice until administration.

3.4.3 In vivo activity-based nanosensor studies

Nanosensors were synthesized by CPC Scientific. ABNs were dosed in mannitol buffer (0.28 M mannitol, 5 mM sodium phosphate monobasic, 15 mM sodium phosphate dibasic, pH 7.0-7.5) and deposited into the lungs by intratracheal instillation (50 µL total volume, 20 µM per ABN). Immediately after dosing, all mice were given a subcutaneous injection of PBS (400 µL) to promote adequate urine volumes for subsequent analysis. For the viral pneumonia models, mice were administered the ABN cocktail 6 days post infection (p.i.). For the bacterial pneumonia models, mice were administered ABNs 16 hours p.i. For all mice, after receiving the ABNs mice were returned to their home cage for one hour with full access to food and water. After this hour their bladder was manually voided, and they were transferred into a urine collection chamber. At the end of the second hour, the bladder was manually voided and the urine was collected, along with any urine that was produced in the collection chamber. The urine samples were then sent to Syneos Health for LC-MS/MS analysis. Reporter quantification by LC-MS/MS was performed as previously described. [56]

3.4.4 Statistical analysis of urine signatures

For disease classification based on urinary activity-based nanosensor signatures, randomly assigned sets of paired data samples consisting of features, i.e., standardized scores of peak area ratio (PAR) of individual urinary reporters measured by LC-MS/MS, and labels, i.e., bacterial or viral, were used to train linear support vector machine classifiers implemented in Python 3. All analyses were run with ten-fold cross validation, and trained classifiers

were tested on randomly assigned, held-out, independent test cohorts. Classification performance was evaluated with receiver operating characteristic (ROC) statistics. Classifier performance was reported as the mean accuracy and AUC across ten-fold independent cross validations.

3.5 Author contributions

M.A. conceived the study design with feedback Sangeeta N. Bhatia. M.A., Henry Ko (H.K.) and Leslie W. Chan (L.W.C.) optimized the CAP models. M.A., L.W.C., and H.K. performed in vivo experiments. M.A. and Ava P Soleimany (A.P.S). analyzed the data. M.A. wrote the text with contributions from A.P.S.

3.6 Acknowledgements

We thank H. Fleming for her thoughtful feedback on the manuscript and T. Samad for her advice.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 4

Engineering nanoparticles for breath-based detection of protease activity

4.1 Introduction

Breath has been a readout for physiological status since the days of Hippocrates, when doctors were trained to detect diseases by smelling patients' breath. [97] However, the precise composition of volatiles found in human breath was first probed in 1971, when Linus Pauling used gas-liquid chromatography to identify around 250 substances. [98] Since then, thousands of endogenously produced volatile organic compounds (VOCs) have been detected among breath, skin, urine, feces, and other bodily tissues, giving rise to the new field of "volatolomics". [99] Breath is a particularly attractive analyte, as it can be sampled repeatedly, non-invasively, at high volumes, and upon request. Furthermore, it contains a diverse set of volatiles that are pooled from around the body due to the constant circulation of blood through the lungs. One known difficulty of using exhaled volatiles as biomarkers for disease is that an individual's volatolome is influenced by a wide range of factors, such as diet, the environment, drugs, and generally anything affecting metabolism. [100] This makes it

particularly difficult to identify VOCs that are differentially enriched in different disease states. Furthermore, individual VOCs, which by definition and nature have low molecular weights, are diluted in the breath, necessitating measurement tools with exquisite levels of sensitivity. [101]

To overcome these limitation, exogenous volatiles, which are VOCs that are not naturally present in the body, can be detected. [101] Early work with exogenous volatiles involved detecting VOCs that were produced by pathogenic bacteria [102], environmental exposures (e.g. the breathalyzer test for ethanol as a readout of blood alcohol concentration), and using isotope labeled volatiles as chemical tracers for metabolic activity. [103] The latter discovery led to the creation of the BreathTek test for *Helicobacter pylori*, a bacterium that colonizes the stomach and is associated with numerous gastrointestinal diseases. Patients are administered an oral solution that contains urea made with ^{13}C, a non-radioactive, naturally occurring carbon isotope. Urease produced by *H. pylori* cleaves the ^{13}C-urea, producing $^{13}CO_2$ that can partition into the breath. If the ratio of $^{13}CO_2/^{12}CO_2$ is above a cutoff ratio, the patient is determined to be *H. pylori* positive. This test demonstrated that breath-based diagnostics can be developed around enzyme activity and used successfully in clinical settings.

Inspired by the use of exogenous volatiles as biomarkers for disease, we sought to modify our our urinary nanoparticles (ABNs) to produce a breath-based readout by replacing the traditional mass encoded reporters with VOCs; we have termed these particles volatile ABNs (vABNs) (Fig. 4-1). The basic structure of these nanoparticles is similar to our traditional ABNs, whereby a PEG scaffold is decorated with peptide substrates. These substrates are linked to a VOC reporter via a peptide bond. Upon substrate cleavage by a target protease, the VOC is released from the nanoparticle and able to partition into the air space of the lung. The VOC is then exhaled, so by collecting breath we are able to use mass spectrometry to measure the reporter concentration. Using this method, we can create a breath signature of disease that is distinct from the signature generated by healthy controls. Throughout the course of this work, the use of probes to produce exogenous VOCs in response to disease

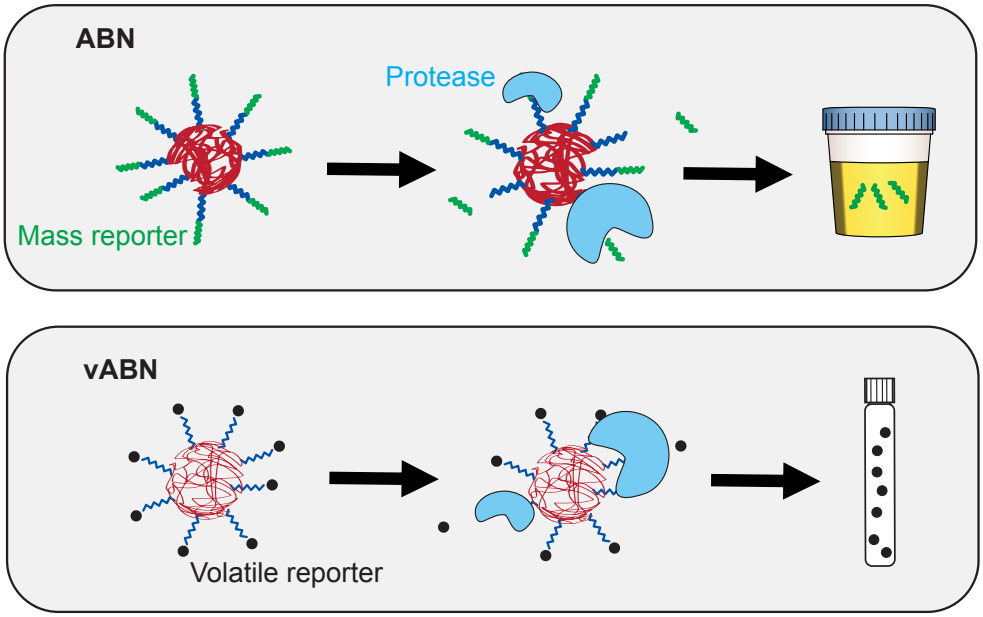


Figure 4-1: Volatile ABNs enable an air-based readout ABNs were modified with volatile organic compound reporters to create volatile ABNs (vABNs). vABNs release their volatile reporters into the airspace, allowing vABN activity to be measured by sampling air.

was termed "induced volatolomics". [99]

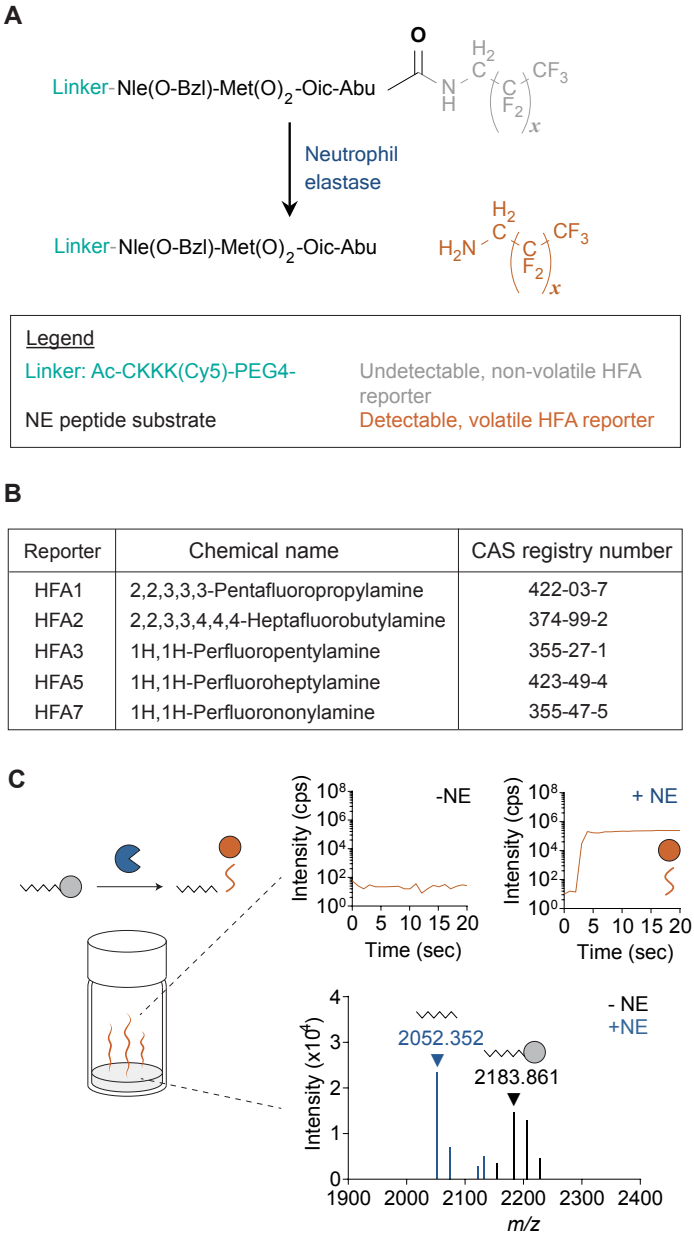
In this work, the serine protease neutrophil elastase (NE) was selected as the target protease. During inflammation, NE is released by neutrophils to kill microbial pathogens and regulate leucocyte recruitment. [104] Due to its elastin-rich composition, the lung is also susceptible to damage by elastases. Therefore, NE plays a notable role in numerous lung diseases, including pulmonary infection, which is the most common risk factor for acute lung injury [105]. Here, we engineer the vABN platform and demonstrate the ability to track pulmonary NE activity through breath analysis within ten minutes of vABN delivery in a mouse pneumonia model with *P. aeruginosa*. We also describe the creation of an *in silico* model for predicting breath signal from these vABNs, and demonstrate that the vABNs can be multiplexed in mouse models of pneumonia virus of mice (PVM) and *S. pneumoniae*.

4.2 Results

4.2.1 Volatile ABNs can release volatile organic compounds in response to protease activity

To engineer vABNs, we chemically modified peptide substrates to release a volatile reporter upon cleavage. We selected an optimized fluorogenic tetrapeptide substrate with the sequence Ac-Nle(O-Bzl)-Met($O)_2$-Oic-Abu-ACC (ACC, fluorophore reporter 7-amino-4-carbamoylmethylcoumarin) due to its selectivity for both human and mouse NE, catalytic efficiency and cleavage site between the C-terminal residue (Abu) and the fluorophore reporter. [106, 107] We hypothesized that an amine-containing VOC could be attached via an amide bond to the peptide C-terminus in place of the fluorophore reporter while maintaining substrate susceptibility to NE. Furthermore, we predicted this attachment chemistry would allow the released VOC to recover its characteristic mass and volatility to undergo phase transition for detection as a gas (Fig. 4-2a).

Hydrofluoroamines (HFAs) with the chemical formula $CF_3(CF_2)_xCH_2NH_2$ were selected as bio-orthogonal VOC reporters due to their high volatility. In this work, HFA reporters will be referred to as HFA_x, where x is the CF_2 chain length (Fig. 4-2b). To test our hypothesis, we synthesized an HFA_1-modified NE peptide substrate and used mass spectrometry to characterize the substrate and reporter behavior in the presence of active NE (Fig. 4-2c). Before NE addition, no HFA_1 was detected in the headspace of vials containing the solubilized, intact substrate. After NE addition, the substrate was cleaved between the Abu residue and HFA_1 reporter, and freed reporters partitioned rapidly from solution into the vial headspace. The HFA_1-releasing substrate was subsequently formulated into vABNs for intrapulmonary delivery. To prolong retention in the lungs, substrates were conjugated to 40 kDa eight-arm poly(ethylene glycol) (PEG) nanocarriers, which distribute throughout the lungs and have a half-life of several days after intratracheal instillation. [56, 108]



Reporter	Chemical name	CAS registry number
HFA1	2,2,3,3,3-Pentafluoropropylamine	422-03-7
HFA2	2,2,3,3,4,4,4-Heptafluorobutylamine	374-99-2
HFA3	1H,1H-Perfluoropentylamine	355-27-1
HFA5	1H,1H-Perfluoroheptylamine	423-49-4
HFA7	1H,1H-Perfluorononylamine	355-47-5

Figure 4-2: vABNs are activated by human NE and release volatile reporters detectable by mass spectrometry. (A) Chemical structure of HFA-modified NE peptide substrates. HFA reporters are attached at the C-terminus via an amide bond. A cysteine is included in the hydrophilic linker at the N-terminus for conjugation to a nanoparticle carrier. (B) Chemical names and CAS registry numbers for volatile reporters. (C) Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrum confirming HFA_1 release after substrate cleavage by purified human NE (bottom) and mass spectra from real-time vapour analysis confirming partitioning of the freed reporter into the reaction headspace (top).

4.2.2 Creation of a mathematical model to predict exhaled vABN concentrations

Having established methods for vABN synthesis, we built a physiologically based pharmacokinetic (PBPK) model to investigate and optimize parameters important for breath signal. Mathematical models have been previously developed to track ABN distribution and predict reporter concentrations in urine [53,109], and PBPK models for the respiratory tract are well-established in the toxicology field. [110–114] Given the novelty of our breath-based readout, we developed a new PBPK model that is governed by a set of differential equations representing the sequence of events from vABN delivery to breath collection. These events include: (i) vABN accumulation in diseased lung tissue, (ii) peptide cleavage by proteases in the lung, and subsequent reporter release, (iii) partitioning of freed reporters from lung tissue to the respiratory lumen, (iv) exhalation of freed reporters and (v) breath analysis (Fig. 4-3). We then used this model to predict the effects of key parameters (i.e. protease concentration, vABN dosing, vABN cleavage kinetics, and reporter partitioning) on reporter concentration in exhaled breath.

Compartment 1: vABN absorption from the lung lumen

In this model, the lung lumen is considered the air space of the lungs and can be represented as the tidal volume. Because we deliver the vABNs into the lungs via intratracheal instillation, we assume that the entire dose is present in the lumen at $t = 0$. Once in the lumen, the vABNs can penetrate into the underlying pulmonary tissue at a rate that we define k_{Tissue}^{NP}.

$$\frac{dC_{lumen}^{NP}}{dt} = -k_{tissue}^{NP}(C_{lumen}^{NP} - C_{tissue}^{NP}) \tag{4.1}$$

Compartment 2: vABN clearance from the pulmonary tissue

In our model, we have grouped the epithelial lining fluid (ELF), the alveolar macrophages within that fluid, and the underlying lung epithelium into a compartment that we refer to as

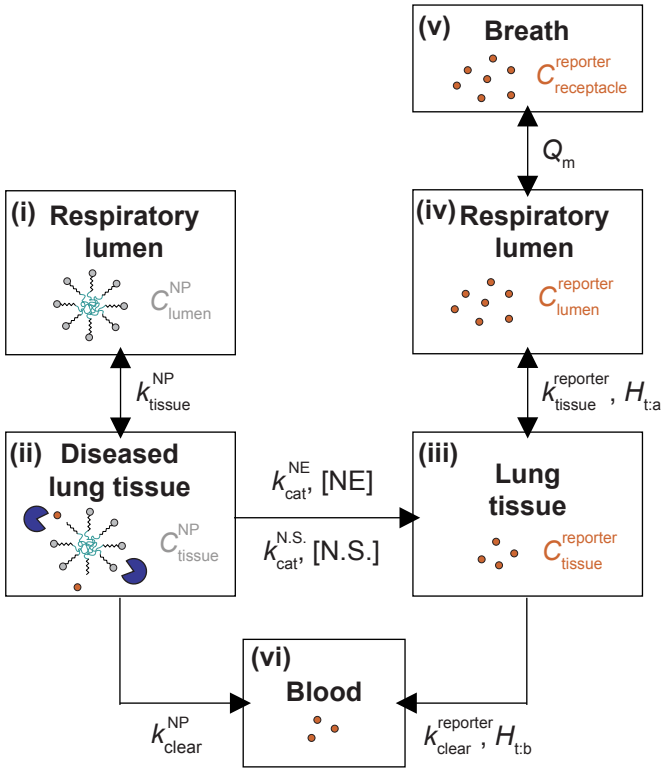


Figure 4-3: PBPK model schematic. Multicompartment PBPK model built to predict volatile reporter concentrations in breath following intrapulmonary vABN delivery (C, concentration; k, rate constant; H, partition coefficient of the free reporter in tissue (t), blood (b), or the respiratory lumen/air (a); Q_m, flow rate). The model consists of differential equations representing the following: diffusion of nanoparticles (NP), or vABNs, from the respiratory lumen into the diseased lung tissue (k^{NP}_{tissue}), generation of freed reporters from NE cleavage of vABNs (k^{NE}_{cat}) and non-specific cleavage by other pulmonary proteases (NS) in the tissue compartment ($k^{N.S.}_{cat}$), partitioning of freed reporters from the tissue into the respiratory lumen ($k^{reporter}_{tissue}$, $H_{t:a}$) and subsequent exhalation in breath (Q_m), and the alternative route of reporter clearance from the lungs by diffusion from tissue into blood ($k^{reporter}_{clear}$, $H_{t:b}$). The model was used to investigate the effects of reporter size on breath signal generated by NE-sensing vABNs during lung infection. Parameters that change with reporter size are $H_{t:b}$ and the catalytic efficiency of vABN cleavage by NE (k_{cat}/K_m).

the "pulmonary tissue". The concentration of the vABNs in the tissue is influenced by cleavage by target proteases (in this description the target is neutrophil elastase, NE), non-specific cleavage by other proteases (non-specific, N.S.), and phagocytosis by alveolar macrophages in the ELF (k_{phago}^{NP}). vABN cleavage is modeled using Michaelis-Menten enzyme kinetics. The kinetic terms describing cleavage by NE (k_{cat}^{NE} and K_m^{NE}) were determined through *in vitro* experiments, while the terms for non-specific enzyme degradation ($k_{cat}^{N.S.}$ and $K_m^{N.S.}$) were fit to *in vivo* data (Fig. 4-4) and assumed to be the same for all vABNs, regardless of their reporter. PEG polymers larger than 20 kDa have been shown to have high retention rates in pulmonary tissue at early timepoints after intratracheal administration. [115] Therefore, because we only modeled the vABN activity up to 3 hours after dosing, our model does not account for transport from the pulmonary tissue into systemic circulation.

$$\frac{dC_{tissue}^{NP}}{dt} = k_{tissue}^{NP}(C_{lumen}^{NP} - C_{tissue}^{NP}) - k_{phago}^{NP}C_{tissue}^{NP} - \frac{(k_{cat}^{NE}[NE]C_{tissue}^{NP}}{(K_m^{NE} + C_{tissue}^{NP})} - \frac{k_{cat}^{N.S.}[N.S.]C_{tissue}^{NP}}{(K_m^{N.S.} + C_{tissue}^{NP})} \quad (4.2)$$

To obtain the concentration of NE in the ELF, we performed bronchoalveolar lavage (BAL) in PA01 infected mice using a single wash (1 mL of phosphate buffer saline instilled, with approximately 80% aspirated). The collected fluid was centrifuged to remove cells, and the supernatant NE concentration was measured via ELISA. To correct for NE dilution in BALF, we normalized this measured concentration, 5.29 ng/mL, to the total volume of ELF in the lungs, which is where NE would be present. Using an ELF volume of 40 μL [116,117], we estimated an NE concentration of 3.5 nM.

$$(5.29ng/mL \text{ of NE in BALF})(0.8mL \text{ of BALF}) = (xng/mL)(0.04mL \text{ of ELF}) \; x = 105.8ng/mL$$

$$105.8ng/mL * (1e3mL)/(1L) * (1g)/(1e9ng) * (1mol)/(29.5e3g) * (1e9nmol)/(1mol) = 3.5nM \text{ NE}$$

Compartment 3: Transport of liberated VOC reporters through the pulmonary tissue

Once the vABNs are cleaved, the liberated reporters can either diffuse from the tissue into the lumen ($k_{tissue}^{reporter}$), which allows them to be exhaled into the chamber, or transported into the blood ($k_{clear}^{reporter}$), after which they are assumed to be diluted in the total circulatory blood volume and unavailable for exhalation.

$$\frac{dC_{tissue}^{reporter}}{dt} = -k_{tissue}^{reporter}(\frac{C_{tissue}^{reporter}}{H_{t:a}} - C_{lumen}^{reporter}) - k_{clear}^{reporter}(\frac{C_{tissue}^{reporter}}{H_{t:b}^{reporter}}) + \frac{k_{cat}^{NE}[NE]C_{tissue}^{NP}}{K_m^{NE} + C_{tissue}^{NP}} + \frac{k_{cat}^{N.S.}[N.S.]C_{tissue}^{NP}}{K_m^{N.S.} + C_{tissue}^{NP}} \quad (4.3)$$

The clearance rate of reporters from the respiratory tissue into blood ($k_{clear}^{reporter}$) was fit to *in vivo* values (Fig. 4-4) and assumed to be equal for all reporters due to their similar molecular weights. $H_{t:a}$ is the tissue:air partition coefficient, which represents the ratio of the VOC reporter concentration in the tissue to the concentration in air. A higher $H_{t:a}$ implies that the reporter partitions more readily into tissue. Similarly, the $H_{b:a}$ represents the ratio of the VOC reporter concentration in the blood to the concentration in air. The $H_{t:b}$ (tissue:blood partition coefficient) was calculated by dividing $H_{t:a}^{reporter}$ by $H_{b:a}^{reporter}$, as is standard for PBPK modeling. The $H_{t:a}^{reporter}$ and $H_{b:a}^{reporter}$ were measured *in vitro* for each reporter except HFA_7 due to difficulties with vaporization, which was assumed to have the same partition coefficients as HFA_5.

Compartment 4: VOC reporter concentration in the pulmonary lumen

The reporters that diffuse from the tissue and into the lung lumen are represented by:

$$\frac{dC_{lumen}^{reporter}}{dt} = k_{tissue}^{reporter}(\frac{C_{tissue}^{reporter}}{H_{t:a}} - C_{lumen}^{reporter}) - Q_{m,c}(C_{lumen}^{reporter} - C_{chamber}^{reporter}) \quad (4.4)$$

The first term of Eq. 4.4 describes the diffusion of the VOC reporter between the lumen and tissue. The second term describes the movement of the reporter that is influenced by breathing. Q_m represents the minute volume, which is the volume of gas that is inhaled or

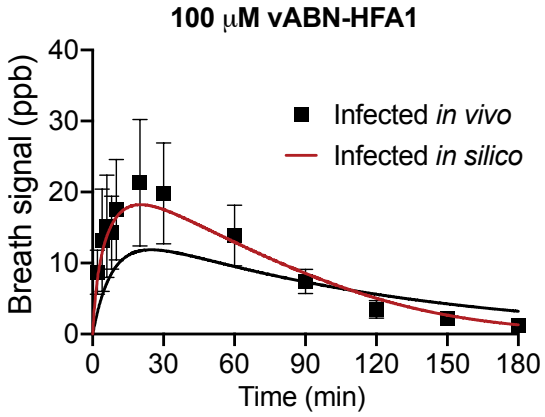


Figure 4-4: In vivo breath signal used to fit the PBPK model. Breath signal after intrapulmonary delivery of HFA_1-releasing vABNs in a mouse model of *P. aeruginosa* was used to fit the PBPK model, in order to determine the values of several unknown parameters. Black squares are mean concentrations from breath measurements.

exhaled per minute. Minute volume consists of both the alveolar volume (the volume of air that undergoes gas exchange) and the dead space (air that is not perfused). The value for Q_m is typically given in units of L/min. To correct for the units, the term $Q_{m,c}$ was used, whereby Q_m is divided by tidal volume (the volume of air displaced between inhalation and exhalation). With this correction, $Q_{m,c}$ effectively represents the ventilation rate.

Compartment 5: The breath collection chamber

Finally, the chamber compartment represents the receptacle that is used to collect breath. In the model, this compartment is the syringe that encloses mice during breath collection. In humans, it could represent any collection container that the patient breathes in and out of during breath collection. The reporters in the lumen that do not diffuse back into the pulmonary tissue are exhaled into the chamber. Assuming ideal transport of VOCs into and out of the lumen, the chamber concentration can be modeled as:

$$\frac{dC_{chamber}^{reporter}}{dt} = Q_{m,c}(C_{lumen}^{reporter} - C_{chamber}^{reporter}) \tag{4.5}$$

Furthermore, this equation accounts for VOC reporters in the chamber that are re-inhaled, which affects the concentration of reporters in the collected breath.

Conversion to parts per billion (ppb)

All concentrations in the model are in terms of μM. To convert the final exhaled concentrations into parts per billion (ppb), the following conversion factor was used:

$$Exh_{ppm} = C_{chamber}^{reporter} * 1e-6 * 1000 * 24450$$
$$Exh_{ppb} = Exh_{ppm} * 1000$$

Parameter estimation

Any values that were not measured or available in the literature were determined by fitting the model for HFA_1 to *in vivo* breath data obtained from administration of 100 μM of the NE responsive HFA_1-vABN into infected mice (Fig. 4-4). The MATLAB function *lsqcurvefit* was used to simultaneously fit $k_{tissue}^{reporter}$, $k_{clear}^{reporter}$, $k_{cat}^{N.S.}$, and $K_m^{N.S.}$ until a local minimum was reached. The fit was bound by the constraints that all values must be positive. k_{tissue}^{NP} was manually adjusted to obtain a value that was less than $k_{tissue}^{reporter}$ and within a range of previously modeled values. [109] The fitted values for $k_{cat}^{N.S.}$ and $K_m^{N.S.}$ were then compared to the measured values of k_{cat}^{NE} and K_m^N for HFA_1, and scaled accordingly to obtain a more generalized form that could be used to account for the differences in cleavage rate for vABNs with other HFA reporters. The $[N.S.]$ was held constant throughout all ABN models, as the concentration of nonspecific enzymes should not change based on the vABN. All parameter values can be found in Table 4.1.

Table 4.1: PBPK Model values. The parameters used to create the PBPK model.

	Parameter	Description	Unit	Value	Source
Flow rates	Q_m	Minute volume	L/min	0.037	[118]
	$Q_{m,c}$	Corrected minute volume	1/min	Q_m/V_{tidal}	Estimate
Volumes	V_{tidal}	Tidal volume	mL	0.131	[118]
Partition coefficients	$H_{b:a}^{reporter}$	VOC reporter blood:air partition coefficient	-	HFA_1: 51.26 HFA_3: 31.50 HFA_5: 18.73	Measurement

Continuation of Table 4.1					
	Parameter	Description	Unit	Value	Source
	$H_{t:a}^{reporter}$	VOC reporter tissue:air partition coefficient	-	HFA_1: 34.49 HFA_3: 36.64 HFA_5: 30.82	Measurement
	$H_{t:b}^{reporter}$	VOC reporter tissue:blood partition coefficient	-	$H_{t:a}^{reporter}/H_{t:b}^{reporter}$	Defined
Clearance rates	k_{tissue}^{NP}	Diffusion rate of vABNs into tissue	1/min	0.05	Manually fit to *in vivo* data
	k_{phago}^{NP}	Clearance rate of vABNs via macrophages	1/min	0.0006	[109]
	$k_{tissue}^{reporter}$	Diffusion rate of VOC into tissue	1/min	30.8	Computationally fit to *in vivo* data
	$k_{clear}^{reporter}$	Clearance rate of VOC into blood	1/min	28.1	Computationally fit to *in vivo* data
Michaelis-Menten Kinetics	[NE]	NE concentration in the respiratory tissue	nM	3.5	Measurement
	k_{cat}^{NE}	Turnover number for NE	1/min	HFA_1: 186 HFA_3: 52.8 HFA_5: 142.2 HFA_7:4.2	Measurement
	K_m^{NE}	Michaelis constant for NE	μM	HFA_1: 10.9 HFA_3: 4.4 HFA_5: 56.1 HFA_7:22.3	Measurement
	[N.S.]	Concentration of nonspecific enzymes in the respiratory tissue	μM	2.62	Computationally fit to *in vivo* data
	$k_{cat}^{N.S.}$	Turnover number for nonspecific enzymes	1/min	$k_{cat}^{NE}/60$	Computationally fit to *in vivo* data
	$K_n^{N.S.}$	Michaelis constant for nonspecific enzymes	μM	$K_m^{NE} * 35$	Computationally fit to *in vivo* data

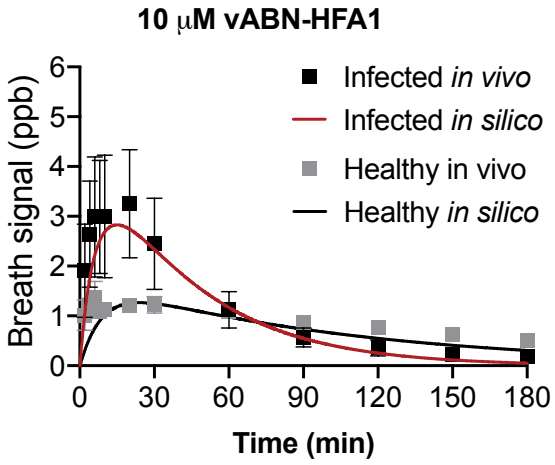


Figure 4-5: Validating the PBPK breath signal predictions. Comparison of *in vivo* breath data to *in silico* predictions for a 10 μM of the NE responsive vABN with the HFA_1 reporter.

4.2.3 Validating the model

Before using the model to make predictions about how changing characteristics of the vABNs and host disease state could influence the breath signal, we first validated that the model was producing accurate predictions by administering a 10 μM dose of the vABN with the HFA_1 reporter into healthy and PA01 infected mice. We found that the model could accurately predict breath signal from both sets of mice, depsite the fact that it was trained on data from a 100 μM vABN dose, assuring us that the parameters could be adjusted while faithfully recapitulating *in vivo* measurements (Fig. 4-5).

4.2.4 The PBPK model can be used to guide vABN design

We hypothesized that protease abundance, vABN dose, cleavage kinetics and partitioning of freed reporters into air (over tissue or blood) were key parameters that would influence the magnitude of the breath signal generated over time. As such, their values were varied in the model to determine their effect on detection sensitivity. Of note, the model predicted that micromolar substrate doses produce ppb-range reporter concentrations in breath, which are well above the mass spectrometer ppt detection limit. Increasing the dose causes the signal to saturate, leading to prolonged breath signal that persists up to two hours after

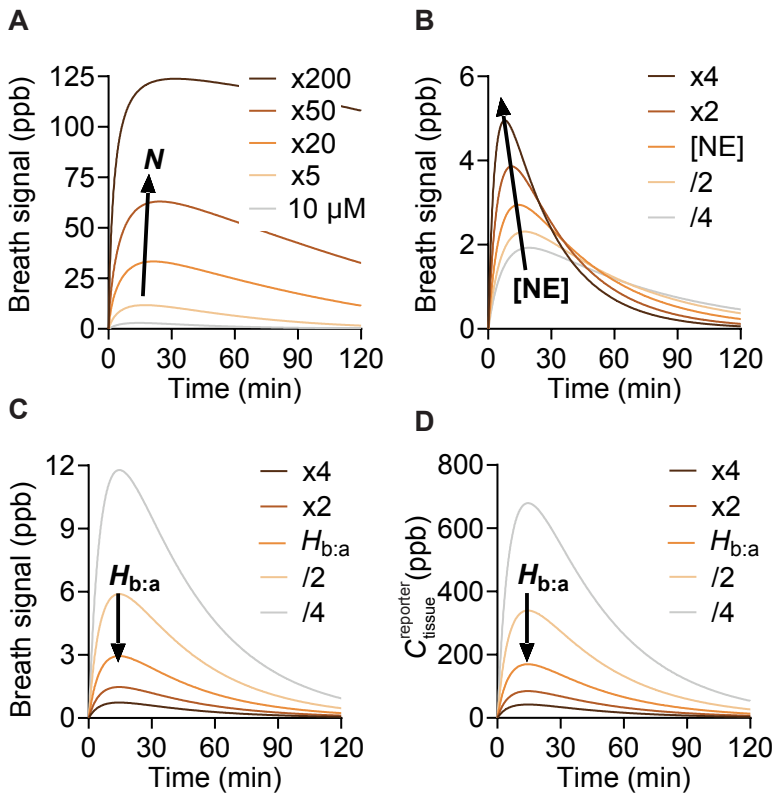


Figure 4-6: Predicted effect of key parameters on breath signal. Key parameters in the PBPK model such as (A) vABN dose (N), (B) NE concentration ([NE]), and (C) blood-air partition coefficient of the reporter ($H_{b:a}$) were varied to confirm model functionality. (A) Micromolar substrate concentration range is predicted to generate breath signal at ppb levels, which is well above the ppt detection limit of mass spectrometry. With increasing vABN dose, we observe increased signal intensity and broadening of signal peaks until a vABN dose is achieved such that even at higher doses, the breath signal remains stable due to establishment of a substrate reservoir. (B) At higher [NE], signal curves are narrowed with higher peak breath signal and earlier return of breath signal to baseline, which can be attributed to faster cleavage of the injected vABN dose. (C,D) As $H_{b:a}$ increases, breath signal drops due to reduced reporter concentrations in tissue (i.e. reporters available to partition into air). Arrows indicate direction of increasing parameter values.

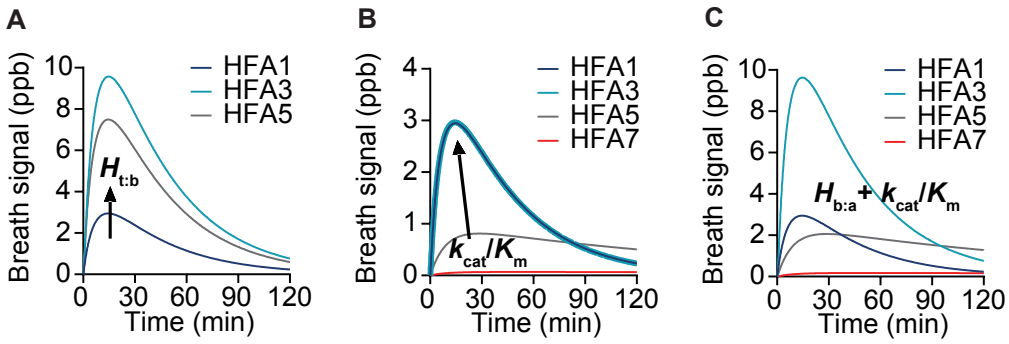


Figure 4-7: In silico investigation of volatile reporter properties. Empirical values for both parameters were determined for a vABN panel containing the candidate reporters (HFA_1, HFA_3, HFA_5 and HFA_7) and used in the model to predict the effect of each parameter in isolation (A,B) and their combined effects (C). Arrows indicate direction of increasing parameter values. Predicted breath signal for HFA_7 is not included in B due to difficulty in vaporizing HFA_7 for determination of its partition coefficient. For d, $H_{t:b}$ for HFA_7 is assumed to be the same as that of HFA_5.

administration (Fig. 4-6a). While such a dose would preclude repeated dosing, it could enable detection with less sensitive instruments. The model also sensibly predicted that increased amounts of NE would shift the peak signal to enable an even faster readout (Fig. 4-6b).

While protease concentrations are inherent to the disease state, and vABN dosing is easily adjusted, other parameters such as cleavage rate and reporter partitioning are governed by vABN composition. vABNs can be synthesized with HFA reporters of varying size, which partition differently into the three compartments — lung tissue (t), blood (b) and respiratory lumen/air (a) — after release. Partition coefficients ($H_{t:a}$ and $H_{b:a}$) and their quotient ($H_{t:b}$) define the ratio of reporters between two compartments at equilibrium. As $H_{b:a}$ increases in favour of the blood compartment, $H_{t:b}$ decreases and more reporter is cleared from lung tissue into blood and unavailable for partitioning into lumen (Fig. 4-6c,d). Therefore, reporters with lower $H_{b:a}$ should provide a greater breath signal.

We determined the partition coefficients for candidate reporters (HFA_1, HFA_3 and HFA_5; Table 4.1) and, using the model, predicted that HFA_3 and HFA_5 would produce the greatest breath signal if partitioning were the only essential parameter (Fig. 4-7a). However, we hypothesized that the benefits of partitioning might be offset by changes to cleavage

rates upon exchanging HFA_1 for larger, bulkier HFA reporters that might adversely affect protease–substrate interactions. To test this hypothesis, we synthesized a vABN panel containing the candidate reporters, and the catalytic efficiency of substrate cleavage (k_{cat}/K_m, where k_{cat} is the catalytic rate constant and K_m is the Michaelis–Menten constant) was determined empirically to be used in the PBPK model (Table 4.1). The catalytic efficiency of HFA_1-containing vABNs fell in between that of the parent, commercial substrate and the optimized peptide substrate (10^3–10^7 $M^{-1}s^{-1}$) [106]. Reduced catalytic efficiency is likely due to nanoparticle sterics [53] and/or the modification of the P1 position [119]. Catalytic efficiency is further reduced by up to two orders of magnitude when HFA_1 is replaced by larger reporters. Thus, when solely considering this parameter, the predicted breath signal was greater for vABNs containing smaller HFAs (Fig. 4-7b). With the exception of HFA_3, which has all-around favourable parameter values, rankings of candidate reporters based on the combined effects of cleavage kinetics and partitioning were nonintuitive. Using the PBPK model, we were able to identify HFA_1 and HFA_3 as our top candidate reporters based on maximum breath signal (Fig. 4-7c).

Using the PBPK model, we explored a final parameter, k_{tissue}^{NP}, that describes the nanoparticle transport rate from the lumen into the lung tissue (that is, the compartment containing NE). In the model, k_{tissue}^{NP} is used to describe the transport of the nanoparticle from the lumen into the underlying lung tissue, the compartment containing the NE. This term represents the combined effects of the permeability constant, surface area, and volume, as defined by Fick's law. When considering how this value would be different for a nanoparticle versus microparticle in the same setting (e.g. the air:tissue interface in the lung), the permeability constant (P) would be the changing variable. $P = \frac{kD}{\Delta X}$, where k is the partition coefficient, D is the diffusion coefficient, and ΔX is the width being traversed. D is dependent on particle size, and using the Stokes-Einstein equation we can estimate the diffusion coefficient in water (at body temperature). For a microparticle with a radius of 0.5 μm and a nanoparticle with a radius of 4 nm (representative of our vABN), D is 5.1e-13 m^2/s and 6.3e-11 m^2/s respectively. Because k_{tissue}^{NP} scales directly with D, this would imply that the k_{tissue}^{NP} for a nanoparticle is about 100 times that of a microparticle. When we model this

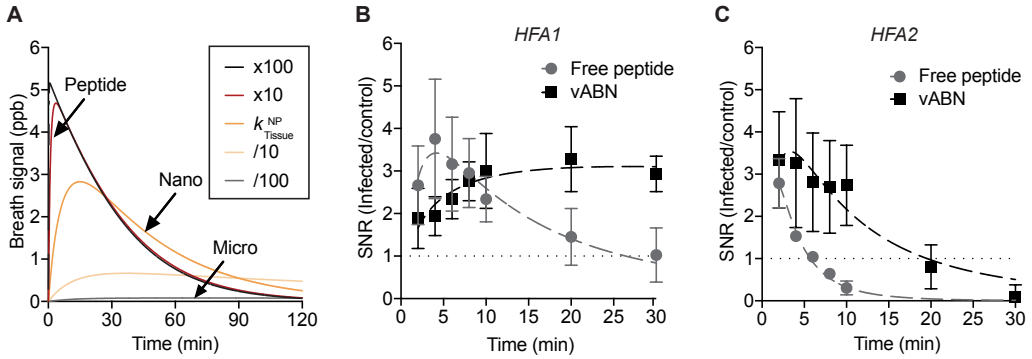


Figure 4-8: Effect of particle size on breath signal. (A) PBPK model predictions for breath signal for free peptide substrates and peptide substrates delivered on nano- or microparticle carriers. (B) Empirical comparison of breath signal after intrapulmonary delivery of the free peptide substrate versus the nanoformulated peptide substrate (i.e. the vABN) containing the HFA_1 reporter (mean ± s.d., n = 5 mice per group). (C) Empirical comparison of breath signal after intrapulmonary delivery of the free peptide substrate versus vABNs containing the HFA_2 reporter (mean ± s.d., n = 4 mice per group).

difference, it becomes clear that a microparticle would not produce a more detectable breath signal, even in the setting of infection (Fig. 4-8a). Conversely, if we increase k_{tissue}^{NP} by an equivalent factor, the model predicts a similar, though delayed, trend to what we observe in the empirical data, such that breath signal decays more rapidly for free peptide substrates than peptide substrates on a nanocarrier.(Fig. 4-8a) We then showed that these kinetics were observed *in vivo* (Fig. 4-8b,c). Ultimately, this modeling justified our nanoformulation over a microformulation or free peptide and with validation from the PBPK model, we moved forward with our nanoscale sensors.

4.2.5 In vivo investigation of vABN derived breath signal

We moved to a mouse model of acute lung infection [67] to verify the feasibility of a protease-driven breath signal (Fig. 4-9a). In this model, NE protein levels in the lungs are elevated 20-fold by 12 hours after intratracheal instillation of Gram-negative *P. aeruginosa* (Fig. 4-9a). After intrapulmonary delivery of the vABNs, we observed breath signal in the exhaled breath as measured by mass spectrometry. Importantly, we confirmed that no background signal is present in the absence of vABN administration (Fig. 4-9b). Repeated sam-

pling revealed that breath collected over 1–3 hours contained dose-dependent, ppb-range reporter concentrations (Fig. 4-9c,d). The maximum breath signal in infected mice was observed 10–20 min after vABN administration, elevated relative to that of healthy controls, and returned to control levels within 1–3 hours. These results are consistent with PBPK model predictions. We next sought to validate model predictions for optimal vABN composition. Breath analysis after administration of 10 μM vABNs containing HFA_1, HFA_3, HFA_5 and HFA_7 reporters revealed that vABNs with HFA_1 and HFA_3 produced the highest breath signal and most rapid signal kinetics (Fig. 4-9e), which was also in line with the PBPK predictions. With these results, we demonstrated that vABNs could produce detectable breath signal that was dose dependent, varied by disease state, and capable of being readout with modular volatile reporters.

4.2.6 vABNs can be multiplexed

As shown in **chapter 3**, one of the strengths of ABNs is the ability to multiplex, as capturing the activity of multiple proteases simultaneously allows for the creation of disease-specific urinary signatures that can be used to train diagnostic classifiers. To determine whether multiplexing could enable breath-based readout of disease etiology, we administered vABNs for Granzyme B (GzmB) and NE into mouse models of PVM and *S. pneumoniae*. The GzmB vABN was synthesized with an HFA_1 reporter, while the NE vABN contained HFA_2. In **chapter 2**, we nominated GzmB and NE as protease biomarkers for viral and bacterial pneumonia, respectively. Therefore, we hypothesized that if we co-administered these vABNs into pneumonia-infected mice, the ratio of HFA_2/HFA_1 would be higher in bacterial pneumonia, due to increased NE activity, and lower in viral pneumonia, due to increased GzmB activity, and vice versa. In a small pilot study, we observed this trend in mice infected with SP (Fig. 4-10b-d) and PVM (Fig. 4-10e-g), thus validating that the vABNs can be multiplexed.

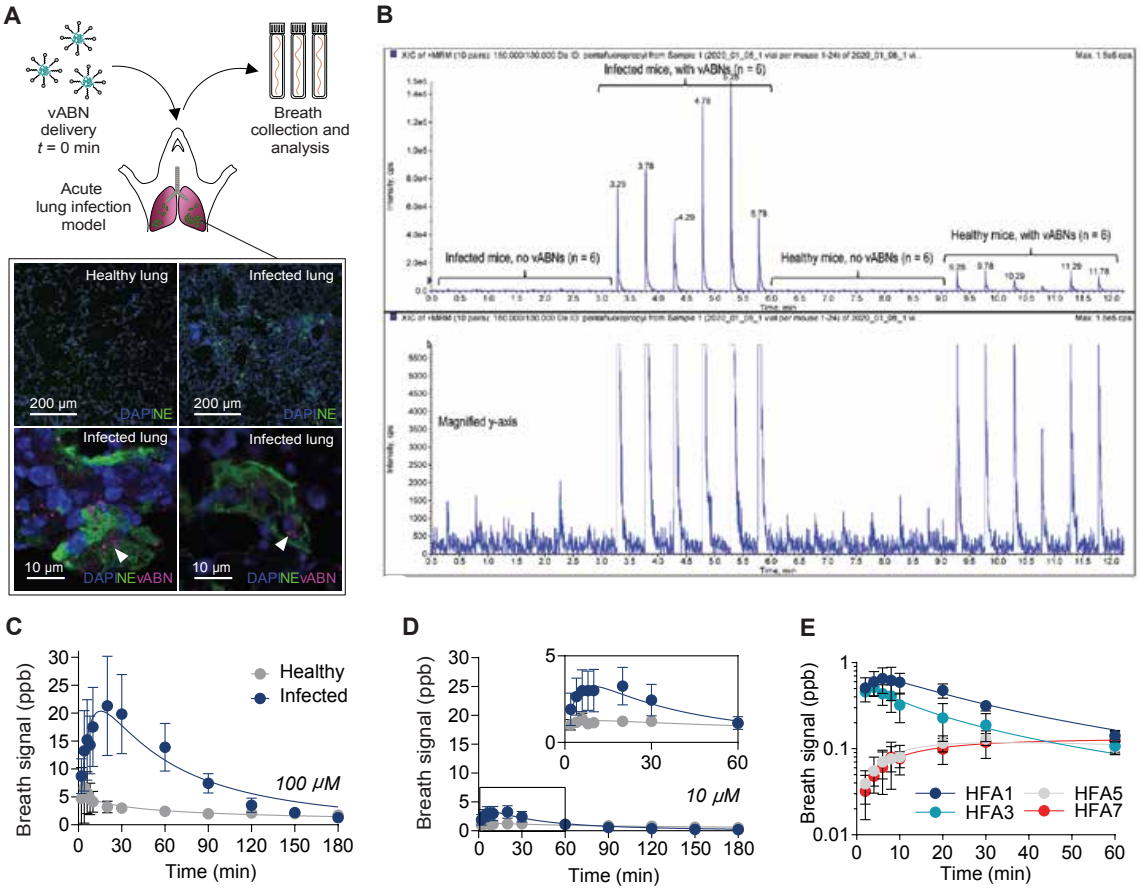


Figure 4-9: Effect of particle size on breath signal. (A) Schematic for characterization of vABN activity in a mouse model of acute lung infection. Immunofluorescence images show NE (green) in healthy lungs (top left) versus lungs infected with P. aeruginosa (top right), and co-localization of vABNs (magenta) with extracellular NE (green, bottom left) and intracellular vABN uptake (bottom right). DAPI staining is used for cell nuclei. White arrows indicate vABNs of interest. Immunofluorescence staining was completed independently twice with similar results. (B) Mass spectra confirming no HFA_1 background signal in the absence of vABN administration. Each peak represents the reporter signal in breath collected from one mouse. Each sample was analyzed by the mass spectrometer for a duration of 0.25 min. (C,D) Breath signal after intrapulmonary delivery of HFA_1-releasing vABNs in healthy controls and lung infection mouse models. Two concentrations were tested, 10 and 100 μM vABN by peptide concentration (mean$\pm$s.d., n=3 or 4 mice per group). Area-under-the-curve analysis showed that the total reporters exhaled scaled with dose (46.1 and 398 pmol HFA_1 for 10 and 100 μM, respectively). (E) Breath signal after delivery of 10 μM vABNs containing HFA_1, HFA_3, HFA_5 or HFA_7 reporters (mean$\pm$s.d., n=3 or 4 mice per group). Breath experiments were completed independently twice with similar results.

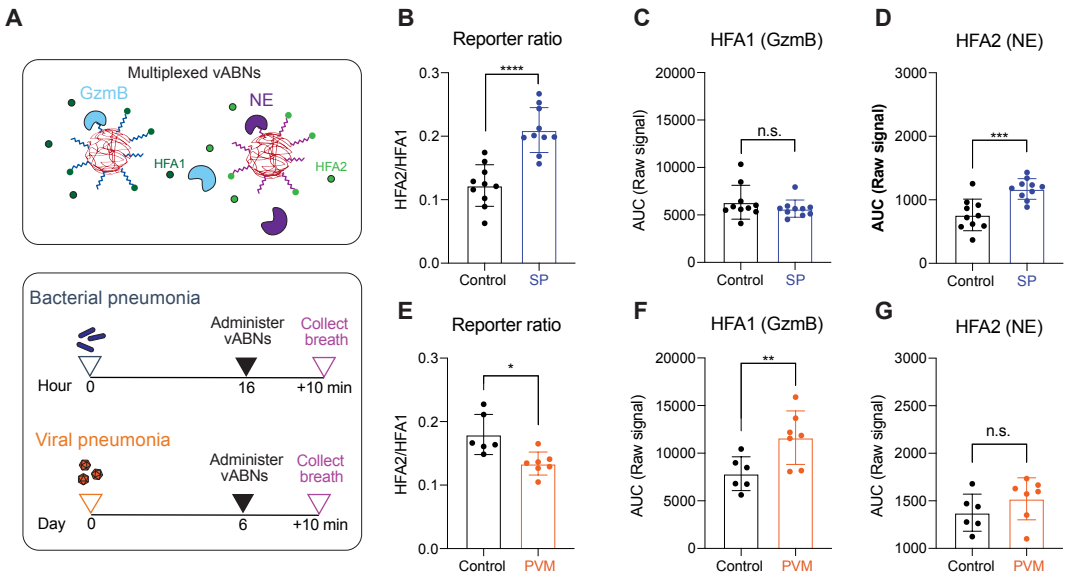


Figure 4-10: Demonstration of multiplexed vABNs for the differentiation of viral vs. bacterial pneumonia. (A) Schematics of the vABN constructs and the experimental timeline. (B) The ratio of the reporter signals from both vABNs after administration into a bacterial pneumonia model. (C-D) Raw values of each HFA reporter as measured by mass spectrometry. (E-G) Corresponding graphs of the reporter ratio and signal from mice infected with viral pneumonia.

4.2.7 Use of the PBPK model to predict breath signal in humans

We can also use our PBPK model for allometric scaling to predict breath signal in humans. The model parameters that differ the most significantly between mice and humans are: (1) the vABN dose, (2) the minute and tidal volumes, (3) the NE concentration in the infected lungs, (4) the concentration of nonspecific enzymes in the respiratory tissue. Based on prior literature, we assume that the partition coefficients hold across species. [120] We assume that the macrophage phagocytosis rate is the same in mice and humans. We can also use the catalytic rate constants (k_{cat}) and Michaelis-Menten constants (K_m) that we obtained *in vitro*, as they were calculated using human NE.

Modification of the vABN dose

Following guidelines outlined by the FDA and literature, we can scale the dose across species by calculating the Human Equivalent Dose (HED) of the vABNs by using Equation 4.6, where k_m (not to be confused with the Michaelis Menten constant, K_m) is an established correction factor for each species that is based on body weight and surface area. [121, 122]

$$\text{HED } (mg/kg) = \text{Animal dose } (mg/kg) * (\text{Animal } k_m/\text{Human } k_m) \tag{4.6}$$

For our animal studies, we administered 10 μM of vABN by peptide concentration in 50 μL of PBS, making our dose in a 30 gram mouse 36.4 $\mu g/kg$. Based on FDA guidelines, the k_m for humans and mice are 37 and 3, respectively. Using Equation 4.6, the human equivalent dose is 2.95 $\mu g/kg$. If we were to deliver this dose by diluting it in 3 mL of saline, which is the saline volume commonly used to administer the adult dose of nebulized albuterol, the starting dose in a 60-kg human would be 27.03 μM. This vABN dose concentration is 2.7 times higher than our current concentration and falls within the concentration range for which acute toxicity was not observed.

Modifying the minute and tidal volumes

The minute volume is the volume of air that is moved in and out of the lungs per minute, while the tidal volume is the volume of air displaced between inhalation and exhalation. [123] The average minute volume for humans is 7.5 L/min, which is 200 times greater than mice (Arms 1988). The tidal volume for a human can be calculated using an established allometric equation $Y = aM^b$, where a and b are established constants specific to the organism (for humans, a = 7.69 and b = 1.04) and M is the mass of the organism in kg. [118] Using this equation, the tidal volume for a 60 kg human is 543.5 mL, which aligns with the 500 mL approximation that is often cited in literature. [123] This is 4000 times greater than mice. Significantly larger human lung volumes could produce greater reporter abundance, but also dilute reporter concentration in breath. The PBPK model factors in this dilution to predict the final breath signal.

NE concentration in an infected human

A prior study found that NE concentration in bronchoalveolar lavage fluid (BALF) from 9 patients with lung infection averaged 780.5 μg/L. [124] BALF was collected by instilling 80-100 mL of saline solution into the lung regions where consolidation was noted on a chest X-ray. Therefore, BALF concentrations are diluted relative to actual concentrations in the epithelial lining fluid (ELF). To estimate the actual NE concentration in ELF (the concentration to which vABNs are exposed), the following figures were used: (1) during BALF collection, average fluid recovery was 47% of the instilled saline solution [124] and (2) in humans, approximately 1 mL of epithelial lining fluid (ELF) is recovered per 100 mL of lavage fluid. [125] Therefore, in 47 mL of lavage fluid, one would expect 0.47 mL of ELF and the estimated NE concentration in human lungs during infection would be the following:

(780.5 $\mu g/L$)(47 mL of BALF) = (x $\mu g/L$)(0.47 mL of ELF)

x = 78.1 mg/L or 2.64 μM of NE in infected human lungs

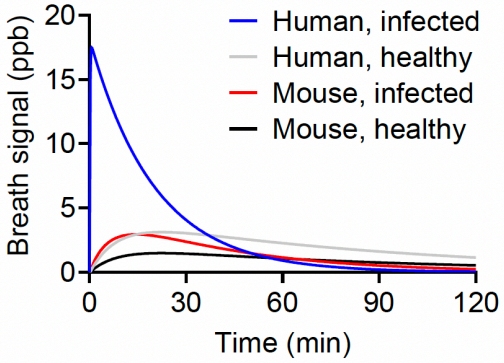


Figure 4-11: Prediction of human breath signal during lung infection using the PBPK model. Human-specific values for parameters such as human equivalent vABN dose, neutrophil elastase concentrations during lung infection, and minute and tidal volumes during respiration were used in the PBPK model to predict breath signal in healthy humans and humans with lung infection. Mouse breath signal predictions are included for comparison.

Concentration of nonspecific (N.S.) enzymes in an infected human

Mouse values for N.S. enzyme concentration were derived by fitting the model to empirical data from breath studies in infected mice. Human values would need to be derived in a similar fashion. For the purposes of our estimation of human breath signal, we assumed an equivalent non-specific enzyme concentration.

Using the adjusted human values, the PBPK model predicts that the breath signal 10 min after vABN administration in humans during lung infection is 4.3 fold higher than in healthy humans and 3.9-fold higher than in infected mice (Fig. 4-11). The absolute signal is at ppb concentrations which is within the limit of detection of the mass spectrometer. These predictions support the feasibility of using our platform to monitor neutrophil elastase activity in humans during lung infection. We expect these values to be even more elevated in during lung infection in individuals with AATD given the deficiency in NE inhibition. Furthermore, collection of human breath onto sorbent tubes, a method in which volatiles in liters of breath can be concentrated onto a solid support for subsequent thermal desorption into a VOC detector, can further increase reporter concentrations up to several orders of magnitude to increase testing sensitivity. [126]

4.3 Discussion

In this work, we have created volatile activity-based nanoparticles, termed vABNs, which release exogenous VOCs into the breath in response to protease activity. By synthesizing vABNs that detect neutrophil elastase (NE) activity, we were able to detect pneumonia via breath in a mouse model of *P. aeruginosa*. We also multiplexed this nanosensor with another vABN responsive to Granzyme B, and showed preliminary data suggesting that multiplexed vABNs can be used to distinguish pneumonia etiology. Finally, we created a mathematical model of vABN activity *in vivo* that can be used to design new nanoparticles and estimate breath signal based on varied *in vitro* and *in vivo* parameters, and showed that the model predicted that our NE vABNs could produce a detectable breath signal in humans. Overall, we have created a proof-of-principle for a breath-based pneumonia diagnostic.

Breath tests are emerging as a simple clinical tool to diagnose disease, as they are convenient, non-invasive, and rapid. There are now a handful of clinical tests that rely on breath and hundreds of clinical trials have involved breath analysis. As a result, there is growing academic, clinical, and industrial interest in breathomics, which puts our work at the forefront of what we expect will be a technological boom. While we have focused on pneumonia as our target disease, breath analysis has been used to detect diseases that touch on every organ system. Given that proteases are also involved in physiological processes underlying a wide range of organs, we expect that our vABNs will be a modular tool that can be used to diagnose this wide range of conditions.

Most excitingly, we believe that by expanding the multiplexing capabilities of our vABN platform, we can translate our urinary 20-plex panel for pneumonia into a breath-based format. This could enable point-of-care detection of pneumonia etiology, as vABNs produce detectable breath signal within 15 minutes rather than 2 hours and portable gas analysis tools exist that could enable detection without the need for expensive and labor-intensive mass spectometers. Being able to quickly determine etiology could enable antibiotic stewardship and ultimately change the diagnostic paradigm for pneumonia. However, there are many difficult challenges that need to be overcome before this vision can become a reality. First, we need to enable multiplexing beyond the capabilities that we have demonstrated.

Based on our urinary results, we would need to create a 5-plex vABN system to diagnose pneumonia and distinguish etiology. This would require the selection of additional VOC reporters that have favorable kinetics and partition properties. Ideally these new reporters would be approved by the FDA, to expedite clinical approval. Second, our vABNs are currently limited by the fact that they require terminal cleavage in order to achieve unscarred release of the VOC reporter. Reconfiguring the chemistry of our substrate linker could allow us to target endopeptidases, which would open up the pool of protease targets dramatically. Finally, we must validate that our vABNs can produce a signal high enough to be detected with portable gas sensing tools, which usually have poorer sensitivity than traditional mass spectrometers. To do this we can use our PBPK model to predict how changes in vABN synthesis and multiplexing would affect breath signal.

4.4 Materials and Methods

4.4.1 vABN synthesis and characterization

HFA-modified peptides were synthesized by CPC Scientific (>95% purity). Briefly, the NE peptide substrate, Ac-CKKK(Cy5)-PEG4-Nle(OBzl)-Met(O)2-Oic-Abu-OH and the GzmB substrate, Ac-CKKK(Cy5)-PEG4-Ile-Glu-Phe-Asp), was synthesized on Fmoc-Abu-CTC resin via standard Fmoc solid phase peptide synthesis. The peptide was cleaved from the resin using 30% HFIP in DCM for 30 min and subsequently coupled to the HFA reporter using DIC/HOBt coupling reagents in DCM under stirring conditions at room temperature for 2h. The finished product was deprotected in TFA for 2.5 hours, precipitated and washed in chilled ether 2 times, and dried under vacuum overnight. The crude peptide was then purified using RP-HPLC. For vABN synthesis, HFA-modified peptides were conjugated to 40 kDa maleimide-functionalized 8-arm PEG (Jenkem) overnight at room temperature in DI water (2:1 molar ratio of peptide to maleimide groups). Excess peptides were removed using spin filters (Millipore, 10 kDa MWCO), and completed vABNs were stored in DI water at 4°C.

4.4.2 Real-time vapor analysis

A triple quadrupole mass spectrometer with 1 L/min sampling rate and ppq to ppt detection limit was used to quantify HFA reporter concentrations in the headspace of *in vitro* reactions and breath samples collected in glass vials during *in vivo* experiments. Multiple reaction monitoring (MRM), in which both [M+H]+ precursor ions and the [M-HF]+ product ions are monitored, was used for sensitive and selective detection of HFA reporters. Data was acquired using Analyst® software.

4.4.3 Partition coefficient studies

Partition coefficients for HFA reporters were derived empirically to be used in the PBPK model. To determine tissue:air partition coefficients ($H_{t:a}$) and blood:air partition coefficients ($H_{b:a}$) of HFA_1, HFA_3, and HFA_5, blood in EDTA solution and lung tissue were collected from female CD-1 mice and aliquoted into 20-mL volatile organic analysis (VOA) glass vials with rubber septa caps (Thermo Scientific™). 5 µL of pure HFAs were vaporized inside separate VOA glass vials and 10 µL of vaporized HFA was transferred to empty reference vials and vials containing EDTA solution, blood, or lung tissue using gastight syringes. Vials were then allowed to equilibrate for 4h at 37°C, after which, 10 μL of headspace from each vial was sampled using a gastight syringe and introduced into the triple quadrupole mass spectrometer. Partition coefficients were calculated using equations established by Gargas et al. [**?**] For $H_{EDTA:a}$ and $H_{t:a}$, where c represents HFA concentration in the headspace, v represents volume, and i represents the compartment of interest (i.e. EDTA or tissue):

$$H_{i:a} = \frac{C_{ref}V_{vial} - C_i(V_{vial} - V_i)}{C_iV_i} \tag{4.7}$$

For the $H_{b:a}$:

$$H_{b:a} = \frac{C_{ref}(V_{vial} - V_{EDTA}) - C_b(V_{vial} - V_{EDTA} - V_b) + (C_{ref} - C_b)H_{EDTA:a}V_{EDTA}}{C_bV_b} \tag{4.8}$$

4.4.4 PBPK Modeling

Code for the PBPK model is available online at `https://github.com/NN19092108A/PBPKmodel`.

4.4.5 Lung infection models

All animal studies were approved by the Massachusetts Institute of Technology's Committee on Animal Care and were completed in accordance with the National Institutes on Health Guide for the Care and Use of Laboratory Animals. For lung infection breath studies, 7- to 8-week-old female CD-1 mice (Charles River) were administered an inoculum of 1.5×10^6 c.f.u. *P. aeruginosa* (strain PA01) in 50 l PBS buffer via intratracheal instillation. The inoculum was prepared by diluting an overnight culture 1:10–1:50 in LB media, allowing the secondary culture to grow to an optical density of approximately 0.5, washing the secondary culture twice with PBS and resuspending the bacteria in PBS. *S. pneumoniae* and pneumonia virus of mice were prepared and administered as previously described.

4.4.6 Breath collection and analysis

For vABN breath tests, 10 μM vABNs in 50 μL PBS were administered into mice via intratracheal instillation. At 10 minutes after vABN administration (unless otherwise specified), mice were placed inside a breath collection apparatus consisting of a 100 mL syringe (Wilburn Medical) connected to a stopcock valve (Cole Parmer, UX-30600-05) with a 23G needle (BD, 305145). The syringe was sealed for 2 minutes to allow breath volatiles to accumulate in the syringe headspace. After 2 minutes, the valve was opened and 55 mL of headspace was displaced into five 12-mL Exetainers (Labco Limited) by puncturing the rubber septum with the needle and pushing the syringe plunger. To measure the HFA reporter signal in breath samples, Exetainers were uncapped at the inlet of a triple quadrupole mass spectrometer and analysed for 0.25 minutes each. Signal peaks were integrated to determine total counts and converted to ppb units.

4.4.7 Statistical analysis

Ordinary one-way ANOVA with Tukey's multiple comparisons test, ROC curve analyses, and calculations for Pearson's correlation coefficient were completed in GraphPad Prism 8.

4.5 Author contributions

Sections of this chapter were derived from [64]. Leslie W. Chan (L.W.C) was the lead author on the paper, and wrote it with contributions from Sangeeta N. Bhatia (S.N.B) and Melodi Anahtar (M.A). The author contributions from that paper are updated below to reflect changes to that original publication that were made for this chapter.

L.W.C. and S.N.B. conceived the study with suggestions from R.R.K. and M.A. L.W.C. and M.A. synthesized and characterized the nanoparticle sensors. L.W.C., M.A., and Ta-Hsuan Ong (T.-H.O.) carried out *in vitro* experiments. M.A. built the multicompartment model for *in silico* predictions of breath signal output, extracted *in vivo* parameters for reporter partitioning and completed *in silico* experiments. L.W.C., M.A., and Kelsey E. Hern (K.E.H.) carried out *in vivo* experiments. L.W.C., M.N.A. and T.-H.O. analysed the data.

Chapter 5

Perspectives and future directions

5.1 Towards the non-invasive diagnosis of pneumonia

Community-acquired pneumonia (CAP) is responsible for significant morbidity and mortality worldwide, and is broadly defined by acute signs and symptoms of lower respiratory infection without other obvious causes. Despite its prevalence, current diagnostics are insufficient to rapidly detect pneumonia and determine etiology. A rapid, accessible pneumonia diagnostic that could simultaneously detect the bacterial and viral pathogens responsible for a patient's CAP would allow doctors to make prompt, fully informed treatment decisions. In this thesis, we sought to create a novel means to diagnose pneumonia and distinguish etiology that is fast, non-invasive, and specific. To this end, we created, screened, tested, and validated intrapulmonary activity-based nanosensors for pneumonia using *in silico*, *in vitro*, *in vivo*, and *in situ* methods. These sensors can create a urine-based readout of etiology within 2 hours and a breath-based readout of pneumonia within 10 minutes, thereby decreasing the gap from presentation to diagnosis. This work sets the stage for further development that might one day translate into a clinically useful diagnostic for the rapid and noninvasive detection of CAP.

5.1.1 Bacterial and viral pneumonia have distinct protease activity profiles

With over 500 human proteases, we hypothesized that the proteolytic host response to infection would be distinct enough to distinguish between bacterial and viral pneumonia. In this work, we demonstrated that this was the case on several fronts. First, we by querying human transcriptomic datasets from blood for proteases that were differentially expressed in bacterial and viral infections, we identified numerous novel pneumonia biomarkers that were associated with distinct biological pathways (Fig. 2-2). We then designed ABNs to target a subset of these proteases and demonstrated that after intratracheal administration, the ABNs produced urinary reporter signatures that were distinct based on disease state (Fig. 2-4). This indicates that there are proteases localized to the lungs that exhibit differential activity based on etiology. By creating an AZP form of one ABN, BV01, which had increased signal in viral compared to bacterial pneumonia, we validated that the signal derived from the ABN was driven by differential Granzyme B activity, due to increased infiltration of the virally infected lung by NKs and CTLs (Fig. 2-8). Together, these results show that despite the similar clinical presentations of bacterial and viral pneumonia, the biological processes underlying etiology are distinct in terms of protease activity.

5.1.2 Machine learning enables powerful integration of multiplexed disease signatures

Our lab has previously demonstrated that multiplexing ABNs is key in distinguishing disease states. [54,56] Whereas these previous studies have shown that multiplexed ABN panels could train binary classifiers with high diagnostic capabilities, our 20-plex panel can be used to train multiclass classifiers that can distinguish between up to six disease states with high accuracy (Fig. 3-3). While down-selection to a 5-plex still enabled three-way classification between bacterial pneumonia, viral pneumonia, and healthy controls with high specificity and sensitivity (Fig. 3-4), the larger ABN panel will likely enable discrimination between pneumonia and other comorbidities or confounding conditions such as pulmonary

hypertension, acute respiratory distress syndrome (ARDS), systemic lupus erythematosus (SLE).

5.1.3 The promise of breath-based diagnostics

Breath is a powerful analyte, as it can be sampled repeatedly, at large volumes, on demand, in many clinical settings, and noninvasively in people of all ages and conditions. By reformulating our ABNs into a breath-based readout, we were able to decrease the time from sensor administration to a diagnostically useful output from by 10-fold (Fig. 4-9). Notably, we multiplexed the vABNs and showed preliminary evidence which suggests that vABNs can distinguish between etiology within 10 minutes of sensor administration. Ideally, a patient would be able to come to their doctor, get tests done, and receive the results within the same visit. Our vABNs make this clinical pipeline possible for diagnosing CAP, and possibly be used to rule out bacterial pneumonia rapidly enough to enable antibiotic stewardship in patients with suspected CAP.

5.2 Future directions

5.2.1 Detecting microbial protease activity

A clear barrier to translation is that there are inherently differences between model organisms and humans. For example, there are likely differences in the cleavage patterns between the mouse and human protease homologs that would necessitate tweaking of the peptide sequences comprising the ABNs for human use. However, our transcriptomic signature was derived from human patients and our *in vitro* screen was performed with recombinant human proteases, so we have confidence that the ABN panel will likely translate into humans, and perhaps perform even better after optimizing the substrates to any human preferred forms. Differences in urinary reporter signatures between model organisms and humans could also be subverted by developing future panels with ABNs that respond to proteases produced by the pathogens themselves, as their expression and subsequent activity would be independent

from the host. For this work we did not target microbial proteases as our goal was to broadly distinguish between etiology, but the inclusion of ABNs that target microbial proteases in future nanosensor panels could enable pathogen-specific detection.

For example, *S. pneumoniae* is the leading cause of bacterial pneumonia. It is commensal in the URT but is an opportunistic pathogen that can cause pneumonia if able to migrate to the LRT. Once in the lungs, its many virulence factors enable it to initiate infection and evade immunity. One of its most prominent factors is Immunoglobulin A1 (IgA1) protease, which cleaves IgA, the main driver of mucosal immunity. This thwarting of mucosal defense pathways allows *S. pneumoniae* to colonize the lungs, due to inhibited opsonization and prevents agglutination. [127, 128] *H. influenzae*, another gram-positive bacterium that is a human restricted pathogen and a leading cause of CAP, also produces IgA proteases. [129] Another proteolytic virulence factor produced by *S. pneumoniae* is High temperature requirement A (HtrA), a serine protease that is conserved across many bacteria and has been implicated as a major factor in *S. pneumoniae* colonization, though its exact substrates and functions are still being studied. Coronaviruses also produce two proteases that are essential for viral replication: 3C-like protease ($3CL^{pro}$) and papain-like protease (PL^{pro}). [130] These proteases cleave the translated viral genome into the nonstructural proteins that enable replication, such as RNA polymerase and helicase. The active sites of these proteases are relatively conserved across coronaviruses, which allowed scientists to leverage previous studies about the role of these protases in MERS and SARS to better understand SARS-CoV-2 when it was first identified.

5.2.2 Detecting additional causes of CAP

An important biological limitation of this work is that a wide range of bacterial and viral pathogens cause CAP. Thus, to use the ABN panel as a point-of-care diagnostic, the classifier will need to be trained on urinary signatures derived from human subjects infected with pathogens beyond the five included in this work. For example, emerging causes of CAP, such as SARS-CoV-2, should be included in classifier training. Furthermore, while fungal CAP is uncommon, it does occur; therefore, the ABNs should also be trained on CAP due

to fungi.

5.2.3 Detecting co-infections and secondary infection

Secondary bacterial infections are known to occur following primary influenza infections. Therefore, further advances in multiplexing could allow the ABNs to be tested in a clinically relevant secondary infection model whereby mice first contract viral pneumonia, regain partial control over the infection, and are subsequently exposed to a bacterial pathogen that causes bacterial pneumonia. Theoretically, as the disease burden shifts from being virally dominated to bacterially driven, the exhaled breath signatures will change accordingly.

5.2.4 Monitoring responses to CAP treatment

Given the repeated sampling that is possible with breath, vABNs could be particularly useful for monitor responses to CAP treatment. This would enable doctors to shorten the duration of a patient's antibiotic regimen. In a hospital setting, this could allow patients to get discharged sooner, thereby reducing their risk of acquiring a secondary nosocomial infection and reducing their medical expenses. In an outpatient setting, it would allow patients to get back to their daily lives and limit their risk of experiencing side effects as a result of their medication.

THIS PAGE INTENTIONALLY LEFT BLANK

Bibliography

[1] Gregory Tsoucalas and Markos Sgantzos. Hippocrates, on the Infection of the Lower Respiratory Tract among the General Population in Ancient Greece. 2016.

[2] Aaron J. Prussin, II, Ellen B. Garcia, and Linsey C. Marr. Total Virus and Bacteria Concentrations in Indoor and Outdoor Air. *Environmental science & technology letters*, 2(4):84, 4 2015.

[3] World Health Organization. The top 10 causes of death, 12 2020.

[4] Joshua P. P. Metlay, Grant W. W. Waterer, Ann C. C. Long, Antonio Anzueto, Jan Brozek, Kristina Crothers, Laura A. A. Cooley, Nathan C. C. Dean, Michael J. J. Fine, Scott A. A. Flanders, Marie R. R. Griffin, Mark L. L. Metersky, Daniel M. M. Musher, Marcos I. I. Restrepo, and Cynthia G. G. Whitney. Diagnosis and treatment of adults with community-acquired pneumonia. *American Journal of Respiratory and Critical Care Medicine*, 200(7):E45–E67, 2019.

[5] Seema Jain, Wesley H. Self, Richard G. Wunderink, Sherene Fakhran, Robert Balk, Anna M. Bramley, Carrie Reed, Carlos G. Grijalva, Evan J. Anderson, D. Mark Courtney, James D. Chappell, Chao Qi, Eric M. Hart, Frank Carroll, Christopher Trabue, Helen K. Donnelly, Derek J. Williams, Yuwei Zhu, Sandra R. Arnold, Krow Ampofo, Grant W. Waterer, Min Levine, Stephen Lindstrom, Jonas M. Winchell, Jacqueline M. Katz, Dean Erdman, Eileen Schneider, Lauri A. Hicks, Jonathan A. McCullers, Andrew T. Pavia, Kathryn M. Edwards, and Lyn Finelli. Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults. *http://dx.doi.org/10.1056/NEJMoa1500245*, 373(5):415–427, 7 2015.

[6] Musher DM, Abers MS, and Bartlett JG. Evolving Understanding of the Causes of Pneumonia in Adults, With Special Attention to the Role of Pneumococcus. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 65(10):1736–1744, 11 2017.

[7] Saeed Shoar and Daniel M Musher. Etiology of community-acquired pneumonia in adults: a systematic review.

[8] Antoni Torres, Catia Cilloniz, Michael S. Niederman, Rosario Menéndez, James D. Chalmers, Richard G. Wunderink, and Tom van der Poll. Pneumonia. *Nature Reviews Disease Primers 2021 7:1*, 7(1):1–28, 4 2021.

[9] J Ferreira-Coimbra, C Sarda, and J Rello. Burden of Community-Acquired Pneumonia and Unmet Clinical Needs. *Advances in therapy*, 37(4):1302–1318, 4 2020.

[10] Michael Klompas. Clinical evaluation and diagnostic testing for community-acquired pneumonia in adults, 9 2021.

[11] Deena Rachel Zimmerman, Nahum Kovalski, and Barak Nahir. Failure to Radiologically Confirm Community-Acquired Pneumonia Means Antibiotic Overtreatment. *Clinical Infectious Diseases*, 54(12):1816–1816, 6 2012.

[12] G Y Shin. What about viral community acquired pneumonias? *BMJ (Clinical research ed.)*, 337(7660):a598, 7 2008.

[13] Alejandro Díaz, Paulina Barria, Michael Niederman, Marcos I. Restrepo, Jorge Dreyse, Gino Fuentes, Bernardita Couble, and Fernando Saldias. Etiology of Community-Acquired Pneumonia in Hospitalized Patients in Chile. *Chest*, 131(3):779–787, 3 2007.

[14] Michael Burk, Karim El-Kersh, Mohamed Saad, Timothy Wiemken, Julio Ramirez, and Rodrigo Cavallazzi. Viral infection in community-acquired pneumonia: a systematic review and meta-analysis. *European Respiratory Review*, 25(140):178 LP – 188, 6 2016.

[15] Andreas E. Zautner, Uwe Groß, Matthias F. Emele, Ralf M. Hagen, and Hagen Frickmann. More pathogenicity or just more pathogens? -On the interpretation problem of multiple pathogen detections with diagnostic multiplex assays, 2017.

[16] Yange Zhang, Senthilkumar K. Sakthivel, Anna Bramley, Seema Jain, Amber Haynes, James D. Chappell, Weston Hymas, Noel Lenny, Anami Patel, Chao Qi, Krow Ampofo, Sandra R. Arnold, Wesley H. Self, Derek J. Williams, David Hillyard, Evan J. Anderson, Carlos G. Grijalva, Yuwei Zhu, Richard G. Wunderink, Kathryn M. Edwards, Andrew T. Pavia, Jonathan A. McCullers, and Dean D. Erdman. Serology enhances molecular diagnosis of respiratory virus infections other than influenza in children and adults hospitalized with community-acquired pneumonia. *Journal of Clinical Microbiology*, 55(1):79–89, 1 2017.

[17] D. S. Hui and P. K.S. Chan. Viral Pneumonia. *Encyclopedia of Respiratory Medicine, Four-Volume Set*, pages 456–466, 7 2018.

[18] Richard R. Watkins and Tracy L. Lemonovich. Diagnosis and Management of Community-Acquired Pneumonia in Adults. *American Family Physician*, 83(11):1299–1306, 6 2011.

[19] Joshua P. Metlay, Grant W. Waterer, Ann C. Long, Antonio Anzueto, Jan Brozek, Kristina Crothers, Laura A. Cooley, Nathan C. Dean, Michael J. Fine, Scott A. Flanders, Marie R. Griffin, Mark L. Metersky, Daniel M. Musher, Marcos I. Restrepo, and Cynthia G. Whitney. Diagnosis and Treatment of Adults with Community-acquired Pneumonia. An Official Clinical Practice Guideline of the American Thoracic Society and Infectious Diseases Society of America. *https://doi.org/10.1164/rccm.201908-1581ST*, 200(7):E45–E67, 10 2019.

[20] Olli Ruuskanen, Elina Lahti, Lance C. Jennings, and David R. Murdoch. Viral pneumonia. *The Lancet*, 377(9773):1264–1275, 2011.

[21] Margaretha C. Minnaard, Alma C. Van De Pol, Joris A.H. De Groot, Niek J. De Wit, Rogier M. Hopstaken, Sanne Van Delft, Herman Goossens, Margareta Ieven, Christine Lammens, Paul Little, Chris C. Butler, Berna D.L. Broekhuizen, and Theo J.M. Verheij. The added diagnostic value of five different C-reactive protein point-of-care test devices in detecting pneumonia in primary care: A nested case-control study. *Scandinavian Journal of Clinical and Laboratory Investigation*, 75(4):291–295, 2015.

[22] Mark B. Pepys and Gideon M. Hirschfield. C-reactive protein: a critical update. *The Journal of Clinical Investigation*, 111(12):1805–1812, 6 2003.

[23] David N. Gilbert. Procalcitonin as a Biomarker in Respiratory Tract Infection. *Clinical Infectious Diseases*, 52(suppl_4):S346–S350, 5 2011.

[24] Liliana Simon, France Gauvin, Devendra K. Amre, Patrick Saint-Louis, and Jacques Lacroix. Serum procalcitonin and C-reactive protein levels as markers of bacterial infection: a systematic review and meta-analysis. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 39(2):206–217, 7 2004.

[25] Josselin Le Bel, Pierre Hausfater, Camille Chenevier-Gobeaux, François-Xavier Blanc, Mikhael Benjoar, Cécile Ficko, Patrick Ray, Christophe Choquet, Xavier Duval, and Yann-Erick Claessens. Diagnostic accuracy of C-reactive protein and procalcitonin in suspected community-acquired pneumonia adults visiting emergency department and having a systematic thoracic CT scan. *Critical Care*, 19, 2015.

[26] David T. Huang, Donald M. Yealy, Michael R. Filbin, Aaron M. Brown, Chung-Chou H. Chang, Yohei Doi, Michael W. Donnino, Jonathan Fine, Michael J. Fine, Michelle A. Fischer, John M. Holst, Peter C. Hou, John A. Kellum, Feras Khan, Michael C. Kurz, Shahram Lotfipour, Frank LoVecchio, Octavia M. Peck-Palmer, Francis Pike, Heather Prunty, Robert L. Sherwin, Lauren Southerland, Thomas Terndrup, Lisa A. Weissfeld, Jonathan Yabes, and Derek C. Angus. Procalcitonin-Guided Use of Antibiotics for Lower Respiratory Tract Infection. *New England Journal of Medicine*, 379(3):236–249, 7 2018.

[27] Emanuela Sozio, Carlo Tascini, Martina Fabris, Federica D'Aurizio, Chiara De Carlo, Elena Graziano, Flavio Bassi, Francesco Sbrana, Andrea Ripoli, Alberto Pagotto, Alessandro Giacinta, Valentina Gerussi, Daniela Visentini, Paola De Stefanis, Maria Merelli, Kordo Saeed, and Francesco Curcio. MR-proADM as prognostic factor of outcome in COVID-19 patients. *Scientific Reports 2021 11:1*, 11(1):1–7, 3 2021.

[28] Timothy E. Sweeney and Purvesh Khatri. Comprehensive Validation of the FAIM3:PLAC8 Ratio in Time-matched Public Gene Expression Data. *https://doi.org/10.1164/rccm.201507-1321LE*, 192(10):1260–1261, 11 2015.

[29] Aleksandra Havelka, Kristina Sejersen, Per Venge, Karlis Pauksens, and Anders Larsson. Calprotectin, a new biomarker for diagnosis of acute respiratory infections. *Scientific Reports 2020 10:1*, 10(1):1–7, 3 2020.

[30] Leber AL, Everhart K, Daly JA, Hopper A, Harrington A, Schreckenberger P, McKinley K, Jones M, Holmberg K, and Kensinger B. Multicenter Evaluation of BioFire FilmArray Respiratory Panel 2 for Detection of Viruses and Bacteria in Nasopharyngeal Swab Samples. *Journal of clinical microbiology*, 56(6), 6 2018.

[31] Barend Mitton, Roxanne Rule, and Mohamed Said. Laboratory evaluation of the BioFire FilmArray Pneumonia plus panel compared to conventional methods for the identification of bacteria in lower respiratory tract specimens: a prospective cross-sectional study from South Africa. *Diagnostic Microbiology and Infectious Disease*, 99(2):115236, 2 2021.

[32] Céline Monard, Jonathan Pehlivan, Gabriel Auger, Sophie Alviset, Alexy Tran Dinh, Paul Duquaire, Nabil Gastli, Camille d'Humières, Adel Maamar, André Boibieux, Marion Baldeyrou, Julien Loubinoux, Olivier Dauwalder, Vincent Cattoir, Laurence Armand-Lefèvre, and Solen Kernéis. Multicenter evaluation of a syndromic rapid multiplex PCR test for early adaptation of antimicrobial therapy in adult patients with pneumonia. *Critical Care 2020 24:1*, 24(1):1–11, 7 2020.

[33] Kaku N, Hashiguchi K, Iwanaga Y, Akamatsu N, Matsuda J, Kosai K, Uno N, Morinaga Y, Kitazaki T, Hasegawa H, Miyazaki T, Fukuda M, Izumikawa K, Mukae H, and Yanagihara K. Evaluation of FilmArray respiratory panel multiplex polymerase chain reaction assay for detection of pathogens in adult outpatients with acute respiratory tract infection. *Journal of infection and chemotherapy : official journal of the Japan Society of Chemotherapy*, 24(9):734–738, 9 2018.

[34] Meera Mehta, Douglas Slain, Lisa Keller, and P Rocco Lasala. The Impact of Biofire Filmarray Respiratory Panel on Antibiotic Usage in the Emergency Department at an Academic Medical Center. *Open Forum Infectious Diseases*, 4(Suppl 1):S354, 10 2017.

[35] Victoria M. Madigan, Vincent G. Sinickas, Dawn Giltrap, Peter Kyriakou, Katherine Ryan, Hiu-Tat Chan, and Vanessa Clifford. Health service impact of testing for respiratory pathogens using cartridge-based multiplex array versus molecular batch testing. *Pathology*, 50(7):758, 12 2018.

[36] Elisa García-Vázquez, María Angeles Marcos, José Mensa, Andrés de Roux, Jordi Puig, Carmen Font, Gloria Francisco, and Antonio Torres. Assessment of the Usefulness of Sputum Culture for Diagnosis of Community-Acquired Pneumonia Using the PORT Predictive Scoring System. *Archives of Internal Medicine*, 164(16):1807–1811, 9 2004.

[37] Lindsay R. Grant, Laura L. Hammitt, David R. Murdoch, Katherine L. O'Brien, and J. Anthony Scott. Procedures for Collection of Induced Sputum Specimens From Children. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 54(Suppl 2):S140, 4 2012.

[38] Byunghyun Kim, Joonghee Kim, You Hwan Jo, Jae Hyuk Lee, Ji Eun Hwang, Min Ji Park, and Sihyung Lee. Prognostic value of pneumococcal urinary antigen test in community-acquired pneumonia. *PLoS ONE*, 13(7), 7 2018.

[39] S Bellew, CG Grijalva, DJ Williams, EJ Anderson, RG Wunderink, Y Zhu, GW Waterer, AM Bramley, S Jain, KM Edwards, and WH Self. Pneumococcal and Legionella Urinary Antigen Tests in Community-acquired Pneumonia: Prospective Evaluation of Indications for Testing. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 68(12):2026–2033, 6 2019.

[40] Catherine Hyams, O. Martin Williams, and Philip Williams. Urinary antigen testing for pneumococcal pneumonia: is there evidence to make its use uncommon in clinical practice? *ERJ Open Research*, 6(1):00223–2019, 1 2020.

[41] Robert P. Dickson, John R. Erb-Downward, Fernando J. Martinez, and Gary B. Huffnagle. The Microbiome and the Respiratory Tract. *http://dx.doi.org/10.1146/annurev-physiol-021115-105238*, 78:481–504, 2 2016.

[42] Lee J. Quinton, Allan J. Walkey, and Joseph P. Mizgerd. Integrative physiology of pneumonia. *Physiological Reviews*, 98(3):1417–1464, 2018.

[43] Gabriel A. Kwong, Sharmistha Ghosh, Lena Gamboa, Christos Patriotis, Sudhir Srivastava, and Sangeeta N. Bhatia. Synthetic biomarkers: a twenty-first century path to early cancer detection, 10 2021.

[44] Mammalian Degradome Database. Human, Mouse and Rat Degradomes in Numbers.

[45] SN Bhatia JS Dudani, AD Warren. Harnessing protease activity to improve cancer care. *Annu. Rev. Cancer Biol.*, 2:353–376, 2017.

[46] Jeffrey A. Whitsett and Theresa Alenghat. Respiratory epithelial cells orchestrate pulmonary innate immunity. *Nature Immunology*, 16(1):27–35, 2015.

[47] Manfred Kopf, Christoph Schneider, and Samuel P. Nobs. The development and function of lung-resident macrophages and dendritic cells. *Nature Immunology*, 16(1):36–44, 2015.

[48] Philippa K Pribul, James Harker, Belinda Wang, Hongwei Wang, John S Tregoning, Jürgen Schwarze, and Peter J M Openshaw. Alveolar Macrophages Are a Major Determinant of Early Responses to Viral Lung Infection but Do Not Influence Subsequent Disease Development. *Journal of Virology*, 82(9):4441, 5 2008.

[49] Takeo Ishii, Raja T. Abboud, Alison M. Wallace, John C. English, Harvey O. Coxson, Richard J. Finley, Karey Shumansky, Peter D. Paré, and Andrew J. Sandford. Alveolar macrophage proteinase/antiproteinase expression in lung function and emphysema. *European Respiratory Journal*, 43(1):82–91, 1 2014.

[50] Alistair J. Standish and Jeffrey N. Weiser. Human Neutrophils Kill Streptococcus pneumoniae via Serine Proteases. *The Journal of Immunology*, 183(4):2602–2609, 8 2009.

[51] Hisanori Domon, Masataka Oda, Tomoki Maekawa, Kosuke Nagai, Wataru Takeda, and Yutaka Terao. Streptococcus pneumoniae disrupts pulmonary immune defence via elastase release following pneumolysin-dependent neutrophil lysis. *Scientific Reports 2016 6:1*, 6(1):1–13, 11 2016.

[52] Christopher Chiu and Peter J. Openshaw. Antiviral B cell and T cell immunity in the lungs. *Nature Immunology*, 16(1):18–26, 2015.

[53] Ester J. Kwon, Jaideep S. Dudani, and Sangeeta N. Bhatia. Ultrasensitive tumour-penetrating nanosensors of protease activity. *Nature Biomedical Engineering*, 1(4), 2017.

[54] Gabriel A Kwong, Geoffrey von Maltzahn, Gayathree Murugappan, Omar Abudayyeh, Steven Mo, Ioannis A Papayannopoulos, Deanna Y Sverdlov, Susan B Liu, Andrew D Warren, Yury Popov, Detlef Schuppan, and Sangeeta N Bhatia. Mass-encoded synthetic biomarkers for multiplexed urinary monitoring of disease. *Nature Biotechnology*, 31(1):63–70, 1 2013.

[55] Jaideep S. Dudani, Maria Ibrahim, Jesse Kirkpatrick, Andrew D. Warren, and Sangeeta N. Bhatia. Classification of prostate cancer using a protease activity nanosensor library. *Proceedings of the National Academy of Sciences of the United States of America*, 115(36):8954–8959, 2018.

[56] Jesse D Kirkpatrick, Andrew D Warren, Ava P Soleimany, Peter M K Westcott, Justin C Voog, Carmen Martin-alonso, Heather E Fleming, and Tuomas Tammela. Urinary detection of lung cancer in mice via noninvasive pulmonary protease profiling. 0262(April), 2020.

[57] JC Holter, F Müller, O Bjørang, HH Samdal, JB Marthinsen, PA Jenum, T Ueland, SS Frøland, P Aukrust, E Husebye, and L Heggelund. Etiology of community-acquired pneumonia and diagnostic yields of microbiological methods: a 3-year prospective study in Norway. *BMC infectious diseases*, 15(1), 2 2015.

[58] Barbara E. Jones, Derrick D. Herman, Charles S. Dela Cruz, Grant W. Waterer, Joshua P. Metlay, Joseph K. Ruminjo, and Carey C. Thomson. Summary for Clinicians: Clinical Practice Guideline for the Diagnosis and Treatment of Community-acquired Pneumonia. *https://doi.org/10.1513/AnnalsATS.201909-704CME*, 17(2):133–138, 1 2020.

[59] Timothy E. Sweeney, Hector R. Wong, and Purvesh Khatri. Robust classification of bacterial and viral infections via integrated host gene expression diagnostics. *Science Translational Medicine*, 8(346), 2016.

[60] Elizabeth Culp and Gerard D Wright. Bacterial proteases, untapped antimicrobial drug targets. *The Journal of Antibiotics 2017 70:4*, 70(4):366–377, 11 2016.

[61] AA Agbowuro, WM Huston, AB Gamble, JDA, and Tyndall. Proteases and protease inhibitors in infectious diseases. *Medicinal research reviews*, 38(4):1295–1331, 7 2018.

[62] Ava P. Soleimany and Sangeeta N. Bhatia. Activity-Based Diagnostics: An Emerging Paradigm for Disease Detection and Monitoring. *Trends in Molecular Medicine*, 26(5):450–468, 2020.

[63] LE Sanman and M Bogyo. Activity-based profiling of proteases. *Annual review of biochemistry*, 83:249–273, 2014.

[64] Leslie W. Chan, Melodi N. Anahtar, Ta Hsuan Ong, Kelsey E. Hern, Roderick R. Kunz, and Sangeeta N. Bhatia. Engineering synthetic breath biomarkers for respiratory disease. *Nature Nanotechnology*, 15(9):792–800, 9 2020.

[65] Kevin Y. Lin, Gabriel A. Kwong, Andrew D. Warren, David K. Wood, and Sangeeta N. Bhatia. Nanoparticles that sense thrombin activity as synthetic urinary biomarkers of thrombosis. *ACS Nano*, 7(10):9001–9009, 2013.

[66] Andrew D. Warren, Shonda T. Gaylord, Kevin C. Ngan, Milena Dumont Milutinovic, Gabriel A. Kwong, Sangeeta N. Bhatia, and David R. Walt. Disease detection by ultrasensitive quantification of microdosed synthetic urinary biomarkers. *Journal of the American Chemical Society*, 136(39):13709–13714, 2014.

[67] Colin G. Buss, Jaideep S. Dudani, Reid T.K. Akana, Heather E. Fleming, and Sangeeta N. Bhatia. Protease activity sensors noninvasively classify bacterial infections and antibiotic responses. *EBioMedicine*, 38:248–256, 2018.

[68] CN Loynachan, AP Soleimany, JS Dudani, Y Lin, A Najer, A Bekdemir, Q Chen, SN Bhatia, and MM Stevens. Renal clearable catalytic gold nanoclusters for in vivo disease monitoring. *Nature nanotechnology*, 14(9):883–890, 9 2019.

[69] TE Sweeney, L Braviak, CM Tato, and P Khatri. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *The Lancet. Respiratory medicine*, 4(3):213–224, 3 2016.

[70] A Kamburov, K Pentchev, H Galicka, C Wierling, H Lehrach, and R Herwig. ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic acids research*, 39(Database issue), 1 2011.

[71] R Herwig, C Hardt, M Lienhard, and A Kamburov. Analyzing and interpreting genome data at the network level with ConsensusPathDB. *Nature protocols*, 11(10):1889–1907, 10 2016.

[72] M Galeas-Pena, N McLaughlin, and D Pociask. The role of the innate immune system on pulmonary infections. *Biological chemistry*, 400(4):443–456, 3 2019.

[73] H Tomlin and AM Piccinini. A complex interplay between the extracellular matrix and the innate immune response to microbial pathogens. *Immunology*, 155(2):186–201, 10 2018.

[74] G N Barber. Host defense, viruses and apoptosis. *Cell Death & Differentiation 2001 8:2*, 8(2):113–126, 4 2001.

[75] Zhiwen Gu and Weifeng Shi. Manipulation of viral infection by deubiquitinating enzymes: New players in host-virus interactions. *Future Microbiology*, 11(11):1435–1446, 11 2016.

[76] MK Isaacson and HL Ploegh. Ubiquitination, ubiquitin-like modifiers, and deubiquitination in viral infection. *Cell host & microbe*, 5(6):559–570, 6 2009.

[77] SJ Lord, RV Rajotte, GS Korbutt, and RC Bleackley. Granzyme B: a natural born killer. *Immunological reviews*, 193:31–38, 6 2003.

[78] Niklas K. Björkström, Benedikt Strunz, and Hans-Gustaf Ljunggren. Natural killer cells in antiviral immunity. *Nature Reviews Immunology 2021*, pages 1–12, 6 2021.

[79] AR Kerr, LA Kirkham, A Kadioglu, PW Andrew, P Garside, H Thompson, and TJ Mitchell. Identification of a detrimental role for NK cells in pneumococcal pneumonia and sepsis in immunocompromised hosts. *Microbes and infection*, 7(5-6):845–852, 5 2005.

[80] LE Carlin, EA Hemann, ZR Zacharias, JW Heusel, and KL Legge. Natural Killer Cell Recruitment to the Lung During Influenza A Virus Infection Is Dependent on CXCR3, CCR5, and Virus Exposure Dose. *Frontiers in immunology*, 9(APR), 4 2018.

[81] Ava P. Soleimany, Jesse D. Kirkpatrick, Susan Su, Jaideep S. Dudani, Qian Zhong, Ahmet Bekdemir, and Sangeeta N. Bhatia. Activatable Zymography Probes Enable In Situ Localization of Protease Dysregulation in Cancer. *Cancer Research*, 81(1):213–224, 1 2021.

[82] R Azeem, D Brimhall, and M Pacula. SAFETY AND TOLERABILITY IN HEALTHY VOLUNTEERS OF THE GLYMPSE BIO TEST SYSTEM-NASH DIAGNOSTIC. . . . , 9 2020.

[83] Andrew D. Warren, Gabriel A. Kwong, David K. Wood, Kevin Y. Lin, and Sangeeta N. Bhatia. Point-of-care diagnostics for noncommunicable diseases using synthetic urinary biomarkers and paper microfluidics. *Proceedings of the National Academy of Sciences*, 111(10):3671–3676, 3 2014.

[84] Liangliang Hao, Renee T. Zhao, Chayanon Ngambenjawong, Heather E. Fleming, and Sangeeta N. Bhatia. CRISPR-Cas-amplified urine biomarkers for multiplexed and portable cancer diagnostics. *bioRxiv*, page 2020.06.17.157180, 6 2020.

[85] Nikolay Kolesnikov, Emma Hastings, Maria Keays, Olga Melnichuk, Y. Amy Tang, Eleanor Williams, Miroslaw Dylag, Natalja Kurbatova, Marco Brandizi, Tony Burdett, Karyn Megy, Ekaterina Pilicheva, Gabriella Rustici, Andrew Tikhonov, Helen Parkinson, Robert Petryszak, Ugis Sarkans, and Alvis Brazma. ArrayExpress update-simplifying data submissions. *Nucleic Acids Research*, 43(D1):D1113–D1116, 1 2015.

[86] Ron Edgar, Michael Domrachev, and Alex E. Lash. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, 30(1):207–210, 1 2002.

[87] W. Evan Johnson, Cheng Li, and Ariel Rabinovic. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, 8(1):118–127, 1 2007.

[88] Virginia Goss Tusher, Robert Tibshirani, and Gilbert Chu. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences*, 98(9):5116–5121, 4 2001.

[89] Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, 1 1995.

[90] C. Patrick Doncaster and Rebecca Spake. Correction for bias in meta-analysis of little-replicated studies. *Methods in Ecology and Evolution*, 9(3):634–644, 3 2018.

[91] Larry V. Hedges. Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, 92(2):490–499, 9 1982.

[92] Larry V. Hedges. A random effects model for effect sizes. *Psychological Bulletin*, 93(2):388–395, 3 1983.

[93] Daniel Lakens. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 0(NOV):863, 2013.

[94] Sylvain Arlot and Alain Celisse. A survey of cross-validation procedures for model selection. *https://doi.org/10.1214/09-SS054*, 4(none):40–79, 1 2010.

[95] M. Stone. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):111–133, 1 1974.

[96] Hong Zheng, Aditya M. Rao, Denis Dermadi, Jiaying Toh, Lara Murphy Jones, Michele Donato, Yiran Liu, Yapeng Su, Cheng L. Dai, Sergey A. Kornilov, Minas Karagiannis, Theodoros Marantos, Yehudit Hasin-Brumshtein, Yudong D. He, Evangelos J. Giamarellos-Bourboulis, James R. Heath, and Purvesh Khatri. Multi-cohort analysis of host immune response identifies conserved protective and detrimental modules associated with severity across viruses. *Immunity*, 54(4):753–768, 4 2021.

[97] The Economist. Smelling bad - Olfactory diagnostics, 2008.

[98] L Pauling, A B Robinson, R Teranishi, and P Cary. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proceedings of the National Academy of Sciences of the United States of America*, 68(10):2374–6, 10 1971.

[99] Fabiola Djago, Justin Lange, and Pauline Poinot. Induced volatolomics of pathologies. *Nature Reviews Chemistry*, 5(3):183–196, 3 2021.

[100] Joachim D. Pleil, Matthew A. Stiegel, and Terence H. Risby. Clinical breath analysis: discriminating between human endogenous compounds and exogenous (environmental) chemical confounders. *Journal of Breath Research*, 7(1):017107, 2 2013.

[101] Edoardo Gaude, Morad K. Nakhleh, Stefano Patassini, Jasper Boschmans, Max Allsworth, Billy Boyle, and Marc P. Van Der Schee. Targeted breath analysis: exogenous volatile organic compounds (EVOC) as metabolic pathway-specific probes. *Journal of Breath Research*, 13(3):032001, 5 2019.

[102] Emma Tait, John D. Perry, Stephen P. Stanforth, and John R. Dean. Use of volatile compounds as a diagnostic tool for the detection of pathogenic bacteria. *TrAC Trends in Analytical Chemistry*, 53:117–125, 1 2014.

[103] P. B. Watkins, S. A. Murray, L. G. Winkelman, D. M. Heuman, S. A. Wrighton, and P. S. Guzelian. Erythromycin breath test as an assay of glucocorticoid-inducible liver cytochromes P-450. Studies in rats and patients. *The Journal of clinical investigation*, 83(2):688–697, 1989.

[104] CTN Pham. Neutrophil serine proteases: specific regulators of inflammation. *Nat. Rev. Immunol.*, 6:541–550, 2006.

[105] GD Rubenfeld. Incidence and outcomes of acute lung injury. *N. Engl. J. Med.*, 353:1685–1693, 2005.

[106] P. Kasperkiewicz, M. Poreba, S. J. Snipas, H. Parker, C. C. Winterbourn, G. S. Salvesen, and M. Drag. Design of ultrasensitive probes for human neutrophil elastase through hybrid combinatorial substrate library profiling. *Proceedings of the National Academy of Sciences*, 111(7):2518–2523, 2014.

[107] Bethany M. Anderson, Daniel P. Poole, Luigi Aurelio, Garrett Z. Ng, Markus Fleischmann, Paulina Kasperkiewicz, Celine Morissette, Marcin Drag, Ian R. van Driel, Brian L. Schmidt, Stephen J. Vanner, Nigel W. Bunnett, and Laura E. Edgington-Mitchell. Application of a chemical probe to detect neutrophil elastase activation during inflammatory bowel disease. *Scientific Reports 2019 9:1*, 9(1):1–12, 9 2019.

[108] Gursahani H, Riggs-Sauthier J, Pfeiffer J, Lechuga-Ballesteros D, and Fishburn CS. Absorption of polyethylene glycol (PEG) polymers: the effect of PEG size on permeability. *Journal of pharmaceutical sciences*, 98(8):2847–2856, 2009.

[109] Gabriel A. Kwong, Jaideep S. Dudani, Emmanuel Carrodeguas, Eric V. Mazumdar, Seyedeh M. Zekavat, and Sangeeta N. Bhatia. Mathematical framework for activity-based cancer biomarkers. *Proceedings of the National Academy of Sciences of the United States of America*, 112(41):12627–12632, 10 2015.

[110] TR Covington C Tan AM Jarabek JG Teeguarden, MS Bogdanffy. A PBPK model for evaluating the impact of aldehyde dehydrogenase polymorphisms on comparative rat and human nasal tissue acetaldehyde dosimetry. *Inhal. Toxicol.*, 20:375–390, 2008.

[111] WA Chiu. Physiologically based pharmacokinetic (PBPK) modeling of interstrain variability in trichloroethylene metabolism in the mouse. *Environ. Health Perspect.*, 122:456–463, 2014.

[112] A Nong. Physiologically based modeling of the inhalation pharmacokinetics of ethylbenzene in B6C3F1 mice. *J. Toxicol. Environ. Heal. Part A*, 70:1838–1848, 2007.

[113] TR Fennell ME Andersen K Krishnan, ML Gargas. A physiologically based description of ethylene oxide dosimetry in the rat. *Toxicol. Ind. Health*, 8:121–140, 1992.

[114] MB Reddy. Inhalation dosimetry modeling with decamethylcyclopentasiloxane in rats and humans. *Toxicol. Sci.*, 105:275–285, 2018.

[115] J Pfeiffer D Lechuga-Ballesteros H Gursahani, J Riggs-Sauthier. Absorption of polyethylene glycol (PEG) polymers: the effect of PEG size on permeability. *J. Pharm. Sci.*, 98:2847–2856, 2009.

[116] Thomas R Korfhagen, Michael D Bruno, Gary F Ross, Karen M Huelsman, Machiko Ikegamit, Alan H Jobet, Susan E Wert, Barry R Stripp, Randal E Morris, Stephan W Glasser, Cindy J Bachurski, Harriet S Iwamoto, Jeffrey A Whitsett, and John A Clements. Altered surfactant function and structure in SP-A gene targeted mice. *Cell Biology*, 93:9594–9599, 1996.

[117] Lin YW, Zhou QT, Cheah SE, Zhao J, Chen K, Wang J, Chan HK, and Li J. Pharmacokinetics/Pharmacodynamics of Pulmonary Delivery of Colistin against Pseudomonas aeruginosa in a Mouse Lung Infection Model. *Antimicrobial agents and chemotherapy*, 61(3), 3 2017.

[118] Michael J Derelanko and Mannfred A Hollinger, editors. *Handbook of Toxicology, Second Edition*. 2 edition, 2002.

[119] AM Strimpler RL Stein. Catalysis by human leukocyte elastase. Aminolysis of acylenzymes by amino acid amides and peptides. *Biochemistry*, 26:2238–2242, 1987.

[120] Aldo Rescigno and Ajit K. Thakur, editors. *New Trends in Pharmacokinetics*. Springer US, 2012.

[121] AnroopB Nair and Shery Jacob. A simple practice guide for dose conversion between animals and human. *Journal of basic and clinical pharmacy*, 7(2):27, 2016.

[122] FDA. Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers. Technical report, 7 2005.

[123] Michael G. Levitzky. *Pulmonary physiology*. McGraw-Hill Education, 9 edition, 2018.

[124] A. Lengas, V. Poletti, L. Pacifico, C. di Domizio, M. Patelli, and L. Spiga. Acute lung inflammation: neutrophil elastase versus neutrophils in the bronchoalveolar lavage–neutrophil elastase reflects better inflammatory intensity. *Intensive care medicine*, 20(5):354–359, 5 1994.

[125] S. I. Rennard, G. Basset, D. Lecossier, K. M. O'Donnell, P. Pinkston, P. G. Martin, and R. G. Crystal. Estimation of volume of epithelial lining fluid recovered by lavage using urea as marker of dilution. *https://doi.org/10.1152/jappl.1986.60.2.532*, 60(2):532–538, 1986.

[126] Oluwasola Lawal, Waqar M. Ahmed, Tamara M.E. Nijsen, Royston Goodacre, and Stephen J. Fowler. Exhaled breath analysis: a review of 'breath-taking' methods for off-line analysis. *Metabolomics : Official journal of the Metabolomic Society*, 13(10), 10 2017.

[127] Zhiming Wang, Jeremy Rahkola, Jasmina S. Redzic, Ying Chih Chi, Norman Tran, Todd Holyoak, Hongjin Zheng, Edward Janoff, and Elan Eisenmesser. Mechanism and inhibition of Streptococcus pneumoniae IgA1 protease. *Nature Communications 2020 11:1*, 11(1):1–8, 11 2020.

[128] Ying Chih Chi, Jeremy T. Rahkola, Agnieszka A. Kendrick, Michael J. Holliday, Natasia Paukovich, Thomas S. Roberts, Edward N. Janoff, and Elan Z. Eisenmesser. Streptococcus pneumoniae IgA1 protease: A metalloprotease that can catalyze in a split manner in vitro. *Protein Science*, 26(3):600–610, 3 2017.

[129] Timothy F. Murphy, Charmaine Kirkham, Megan M. Jones, Sanjay Sethi, Yong Kong, and Melinda M. Pettigrew. Expression of IgA Proteases by Haemophilus influenzae in the Respiratory Tract of Adults With Chronic Obstructive Pulmonary Disease. *The Journal of Infectious Diseases*, 212(11):1798, 2015.

[130] Chaima Mouffouk, Soumia Mouffouk, Sara Mouffouk, Leila Hambaba, and Hamada Haba. Flavonols as potential antiviral drugs targeting SARS-CoV-2 proteases (3CLpro and PLpro), spike protein, RNA-dependent RNA polymerase (RdRp) and angiotensin-converting enzyme II receptor (ACE2). *European Journal of Pharmacology*, 891:173759, 1 2021.