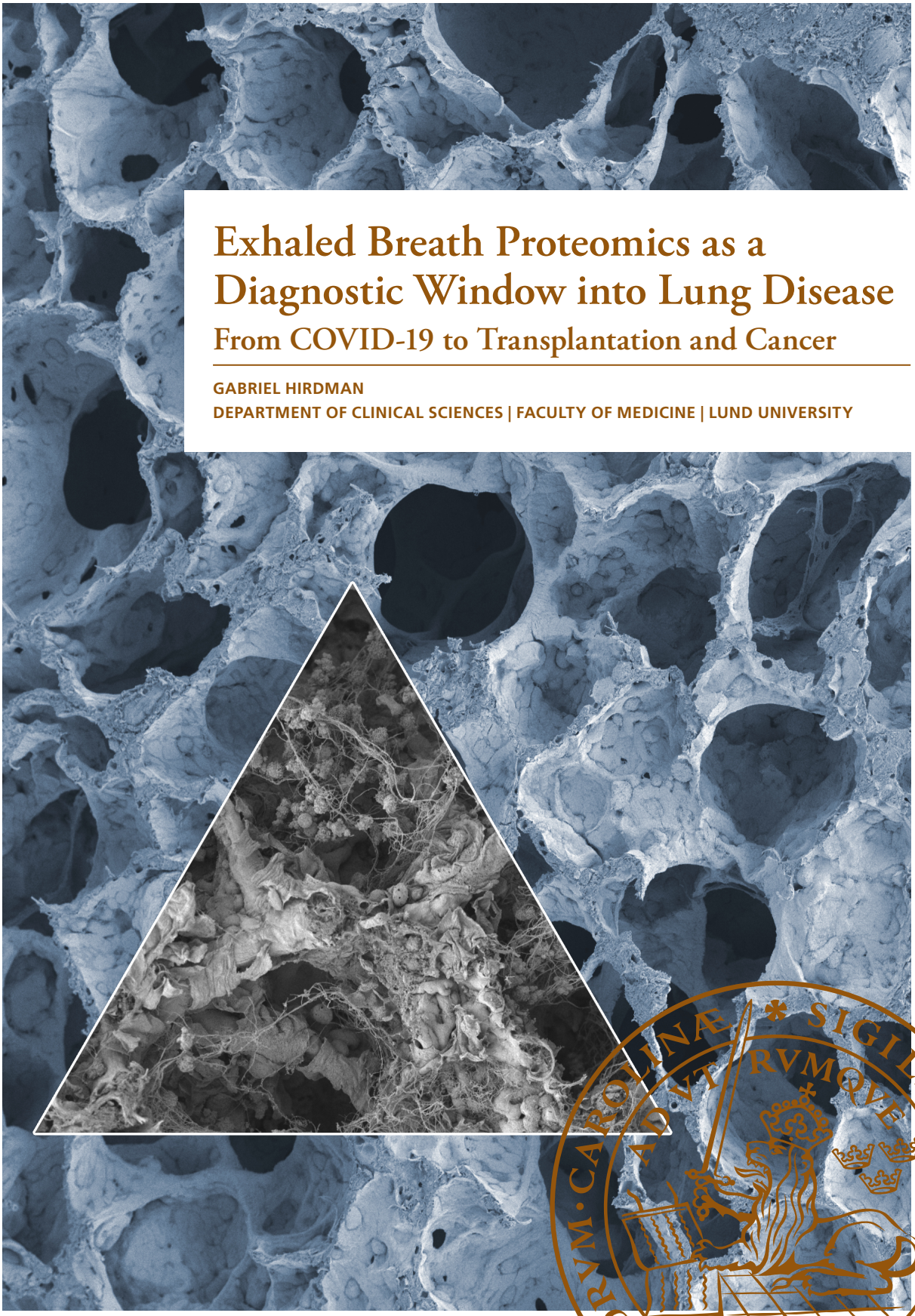
Exhaled Breath Proteomics as a Diagnostic Window into Lund Disease: From COVID-19 to Transplantation and Cancer

Hirdman, Gabriel

2026

Document Version:
Publisher's PDF, also known as Version of record



Exhaled Breath Proteomics as a Diagnostic Window into Lung Disease

From COVID-19 to Transplantation and Cancer

GABRIEL HIRDMAN
DEPARTMENT OF CLINICAL SCIENCES | FACULTY OF MEDICINE | LUND UNIVERSITY

Exhaled Breath Proteomics as a Diagnostic Window into Lung Disease

Exhaled Breath Proteomics as a Diagnostic Window into Lung Disease

From COVID-19 to Transplantation and Cancer

Gabriel Hirdman, MD

LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Medicine at Lund University to be publicly defended on the 25th of May at 08.30 in Belfragesalen, BMC D15, Lund

Faculty opponent
Professor Olav Rooyackers, MD, PhD
Karolinska Institutet, Stockholm, Sweden

Organization: LUND UNIVERSITY

Document name: Doctoral Dissertation Series 2026:85 **Date of issue:** 2026-03-25

Author(s): Gabriel Hirdman, MD **Sponsoring organization:**

Title and subtitle: Exhaled Breath Proteomics as a Diagnostic Window into Lung Disease: From COVID-19 to Transplantation and Cancer

Abstract:

Introduction: Current diagnostic techniques for lung disease are invasive, lack sensitivity, or expose patients to radiation. Exhaled breath particles (EBP) present a novel, non-invasive method for sampling the respiratory tract lining fluid (RTLF) from the distal airways and provide a snapshot of the lungs' proteomic milieu.

Methods: The proteomic composition and diagnostic potential of EBP were evaluated in conjunction with tissue, plasma, and bronchoalveolar lavage fluid (BALF) samples using liquid chromatography-tandem mass spectrometry (LC-MS/MS) across four human and porcine studies. Study cohorts included patients with COVID-19, porcine and human lung transplant recipients with primary graft dysfunction (PGD), porcine models of donor lung injury, and patients with non-small cell lung cancer (NSCLC). Particle concentrations were recorded, and bioinformatic analyses, including machine learning classifiers, were applied to the LC-MS/MS data.

Results: Particle flow rates (PFR) were significantly increased in patients with acute respiratory symptoms and correlated strongly with alveolar-capillary barrier breakdown in PGD and lung injury. Bioinformatic analysis revealed distinct proteomic signatures of lung disease. Acute-phase proteins and surfactant were significantly dysregulated in the EBP of COVID-19 patients. In donor lung injury models, extracellular matrix proteins, particularly collagen VI, were consistently degraded. The EBP proteome of NSCLC was highly enriched for extracellular vesicles and neutrophil extracellular traps. Machine learning classifiers trained on EBP proteomics achieved a diagnostic accuracy of 92% for COVID-19 and 80% for NSCLC samples.

Conclusion: EBP sampling with LC-MS/MS analysis represents a safe and effective non-invasive modality for sampling the distal airway proteome. Proteomic profiling of EBP provides insights into lung pathologies and offers potential for early diagnosis and monitoring of acute and malignant pulmonary disease.

Key words: Exhaled breath analysis, biomarkers, COVID-19, lung transplantation, lung cancer, LC-MS/MS

Language: English, Swedish **Number of pages**: 100

ISSN and key title: 1652-8220, 2026:85

ISBN: 978-91-8021-883-2

Signature **Date** 2026-03-25

Exhaled Breath Proteomics as a Diagnostic Window into Lung Disease

From COVID-19 to Transplantation and Cancer

Gabriel Hirdman, MD

LUND
UNIVERSITY

Supervisor: Professor Sandra Lindstedt, MD, PhD

Co-supervisor: Franziska Olm, PhD

Co-supervisor: Sven Kjellström, PhD

Lund University, Faculty of Medicine

Department of Clinical Sciences

ISBN 978-91-8021-883-2 (print)

Series title: Lund University, Faculty of Medicine Doctoral Dissertation Series 2026:85

ISSN: 1652-8220

Printed in Sweden by Media-Tryck, Lund University,
Lund, 2026

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

To Else

Table of Contents

Abstract

Introduction: Current diagnostic techniques for lung disease are invasive, lack sensitivity, or expose patients to radiation. Exhaled breath particles (EBP) present a novel, non-invasive method for sampling the respiratory tract lining fluid (RTLF) from the distal airways and provide a snapshot of the lungs' proteomic milieu.

Methods: The proteomic composition and diagnostic potential of EBP were evaluated in conjunction with tissue, plasma, and bronchoalveolar lavage fluid (BALF) samples using liquid chromatography-tandem mass spectrometry (LC-MS/MS) across four human and porcine studies. Study cohorts included patients with COVID-19, porcine and human lung transplant recipients with primary graft dysfunction (PGD), porcine models of donor lung injury, and patients with non-small cell lung cancer (NSCLC). Particle concentrations were recorded, and bioinformatic analyses, including machine learning classifiers, were applied to the LC-MS/MS data.

Results: Particle flow rates (PFR) were significantly increased in patients with acute respiratory symptoms and correlated strongly with alveolar-capillary barrier breakdown in PGD and lung injury. Bioinformatic analysis revealed distinct proteomic signatures of lung disease. Acute-phase proteins and surfactant were significantly dysregulated in the EBP of COVID-19 patients. In donor lung injury models, extracellular matrix proteins, particularly collagen VI, were consistently degraded. The EBP proteome of NSCLC was highly enriched for extracellular vesicles and neutrophil extracellular traps. Machine learning classifiers trained on EBP proteomics achieved a diagnostic accuracy of 92% for COVID-19 and 80% for NSCLC samples.

Conclusion: EBP sampling with LC-MS/MS analysis represents a safe and effective non-invasive modality for sampling the distal airway proteome. Proteomic profiling of EBP provides insights into lung pathologies and offers potential for early diagnosis and monitoring of acute and malignant pulmonary disease.

Populärvetenskaplig sammanfattning

Våra lungor består av ett komplext system av celler vars huvudsakliga uppgift är att upprätthålla liv genom att utväxla kroppsproducerad koldioxid mot livsviktigt syre, som ger kroppens celler energi. Om vi skulle sprida ut våra lungor platt skulle dess yta mer eller mindre täcka en vanlig badmintonbana. Över denna yta av tunn lungvävnad återfinns ett tunt skikt av vätska som skyddar våra lungor mot omvärlden. Dagligen andas vi in cirka 10 000 liter luft, innehållande allt från föroreningar och virus till bakterier och cigarettrök. Den tunna vätskan innehåller därför allt från antikroppar till proteiner med uppgift att ta hand om dessa skadliga ämnen. Proteiner kan ses som små byggarbetare i kroppen med olika uppgifter och kan göra allt från att oskadliggöra inkräktare till att skydda de känsliga lungcellerna och användas för kommunikation dem emellan. Vid lungsjukdom, vare sig det rör sig om en covid-19-infektion, avstötning av en transplanterad lunga eller lungcancer, återspeglas detta i sammansättningen av proteiner och antikroppar i den skyddande vätskan. Genom århundradena har denna vätska varit närmast oåtkomlig för läkare och endast kunnat nås genom invasiva undersökningar, såsom lungsköljningar och vävnadsprover som utförs med luftrörskamera.

Varje gång vi andas ut riktigt djupt och fullständigt tömmer lungorna på luft kollapsar de allra djupaste delarna av lungorna. När vi sedan tar ett djupt andetag igen öppnas dessa delar, och små partiklar av den skyddande lungvätskan följer med ut. Med hjälp av en särskild maskin kan vi sedan samla in dessa partiklar, mäta antalet i realtid och därefter undersöka deras sammansättning. Analysen utförs med en extremt känslig typ av mätinstrument som kan identifiera tusentals olika proteiner ur ett enda andetagsprov. Slutresultatet blir som ett blodprov, men provet kommer istället direkt från lungan och späds inte ut med proteiner från andra delar av kroppen. I denna avhandling undersöker vi sammansättningen av dessa proteiner i partiklarna från vätskan vid olika typer av lungsjukdomar.

Vi började med att undersöka covid-19-patienter under coronapandemin och fann att de andades ut en kraftigt ökad mängd partiklar jämfört med friska individer. När vi sedan undersökte nivåerna av olika proteiner hos patienterna fann vi att de hade ökade mängder av just inflammatoriska proteiner och antikroppar. Genom att använda oss av AI kunde vi därefter skilja utandningsluften från friska och sjuka individer åt med över 90 procents säkerhet. I den andra studien undersökte vi utandningsluften efter lungtransplantation och fann återigen att antalet utandade partiklar steg kraftigt hos de patienter vars lungor tidigt började svikta. Detta kunde vi mäta även när patienten behövde konstgjord andning, vilket ofta krävs de första dagarna efter en lungtransplantation, genom att helt enkelt koppla in partikeluppsamlaren i andningsmaskinen. När lungorna sviktar kort efter en lungtransplantation har man sett att det ofta beror på ett ökat vätskeläckage ut i lungorna från blodkärlen, vilket vi kunde koppla till den ökade mängden partiklar

och metoden skulle därför kunna användas för att varna transplantationskirurger tidigt när den transplanterade lungan behöver stöd.

Vi undersökte sedermera även olika typer av lungskador som kan uppstå vid organdonation av lungor. Dessa inkluderar infektionsskador, inhalation av magsaft som fräter på lungorna samt skadlig konstgjord andning. Vi fann att alla dessa typer av lungskador skadade lungornas "byggnadsställning", även kallad det extracellulära matrixet, på likartade sätt och ledde till en ökad mängd partiklar från lungorna. Partiklarna innehöll även här information som kunde skvallra om varför lungskadan uppkommit, såsom magsaftsproteiner. Slutligen, i vår största studie, omfattande över 200 patienter med och utan lungcancer, fann vi kända proteinförändringar förknippade med lungcancer. Genom att återigen använda oss av AI kunde vi med cirka 80 procents säkerhet avgöra vilka av patienterna som hade lungcancer enbart baserat på deras utandningsluft.

Genom att använda oss av nya, känsliga metoder för att detektera proteiner och artificiell intelligens visar denna avhandling att ett djupt andetag kan ge oss en bild av lungornas hälsa och varna vid lungsjukdom. Genom att i framtiden utveckla dessa "fingeravtryck" av proteinsammansättning för olika typer av lungsjukdom är förhoppningen att patienter ska kunna lämna ett utandningsprov och få en diagnos utan att behöva varken nålstick, strålning från röntgen eller vävnadsprov.

List of Publications

List of papers included in the thesis

Study I - **Hirdman, G**., Bodén, E., Kjellström, S., Fraenkel, C.-J., Olm, F., Hallgren, O., & Lindstedt, S. (2023). Proteomic characteristics and diagnostic potential of exhaled breath particles in patients with COVID-19. Clinical Proteomics, 20(1), Artikel 13. https://doi.org/10.1186/s12014-023-09403-2

Study II - Niroomand, A., **Hirdman, G**., Bèchet, N., Ghaidan, H., Stenlo, M., Kjellström, S., Isaksson, M., Broberg, E., Pierre, L., Hyllén, S., Olm, F., & Lindstedt, S. (2024). Proteomic Analysis of Primary Graft Dysfunction in Porcine Lung Transplantation Reveals Alveolar-Capillary Barrier Changes Underlying the High Particle Flow Rate in Exhaled Breath. Transplant International, 37, Artikel 12298. https://doi.org/10.3389/ti.2024.12298

Study III - **Hirdman, G**., Stenlo, M., Bèchet, N. B., Niroomand, A., Mittendorfer, M., Wang, Q., Edström, D., Ghaidan, H., Kjellström, S., Pierre, L., Olm, F., Hyllén, S., & Lindstedt, S. (2025). Unraveling Molecular and Functional Responses Across 3 Lung Injury Models to Expand the Donor Lung Pool. Transplantation, 109(7), 1166-1174. https://doi.org/10.1097/TP.0000000000005353

Study IV - **Hirdman, G**., Andreasson, J., Ghaidan, H., Bodén, E., Planck, M., Hallgren, O., Olm, F., Kjellström, S., & Lindstedt, S. (2026). Identification of Lung Cancer Proteomic Biomarkers in Exhaled Breath Particles: Implications for Non-invasive Diagnosis. (Manuscript).

List of papers outside the thesis scope

Niroomand, A., **Hirdman, G**., Olm, F., & Lindstedt, S. (2022). Current Status and Future Perspectives on Machine Perfusion: A Treatment Platform to Restore and Regenerate Injured Lungs Using Cell and Cytokine Adsorption Therapy. Cells, 11(1), Artikel 91. https://doi.org/10.3390/cells11010091

Ghaidan, H., Stenlo, M., Niroomand, A., Mittendorfer, M., **Hirdman, G**., Gvazava, N., Edström, D., Augusto Silva, I., Broberg, E., Hallgren, O., Olm, F., Wagner, D. E., Pierre, L., Hyllén, S., & Lindstedt, S. (2022). Reduction of primary graft dysfunction using cytokine adsorption during organ preservation and after lung transplantation. Nature Communications, 13(1), Artikel 4173. https://doi.org/10.1038/s41467-022-31811-5

Hirdman, G., Niroomand, A., Olm, F., & Lindstedt, S. (2022). Taking a Deep Breath: an Examination of Current Controversies in Surgical Procedures in Lung Transplantation. Current transplantation reports, 9(3), 160–172. https://doi.org/10.1007/s40472-022-00367-0

Bodén, E., Andreasson, J., **Hirdman, G**., Malmsjö, M., & Lindstedt, S. (2022). Quantitative Proteomics Indicate Radical Removal of Non-Small Cell Lung Cancer and Predict Outcome. Biomedicines, 10(11), Artikel 2738. https://doi.org/10.3390/biomedicines10112738

Niroomand, A., **Hirdman, G**., Pierre, L., Ghaidan, H., Kjellström, S., Stenlo, M., Hyllén, S., Olm, F., & Lindstedt, S. (2023). Proteomic changes to immune and inflammatory processes underlie lung preservation using ex vivo cytokine adsorption. Frontiers in Cardiovascular Medicine, 10, Artikel 1274444. https://doi.org/10.3389/fcvm.2023.1274444

Lindstedt, S., Niroomand, A., Mittendorfer, M., **Hirdman, G.,** Hyllén, S., Pierre, L., & Olm, F. (2023). Nothing but NETs: Cytokine adsorption correlates with lower circulating nucleosomes and is associated with decreased primary graft dysfunction. The Journal of heart and lung transplantation: the official publication of the International Society for Heart Transplantation, 42(10), 1358-1362. https://doi.org/10.1016/j.healun.2023.06.011

Edström, D., Niroomand, A., Stenlo, M., Uvebrant, K., Bölükbas, D. A., **Hirdman, G**., Broberg, E., Lim, H. C., Hyllén, S., Lundgren-Åkerlund, E., Pierre, L., Olm, F., & Lindstedt, S. (2023). Integrin α10β1-selected mesenchymal stem cells reduced hypercoagulopathy in a porcine model of acute respiratory distress

syndrome. Respiratory Research, 24(1), Artikel 145. https://doi.org/10.1186/s12931-023-02459-6

Edström, D., Niroomand, A., Stenlo, M., Broberg, E., **Hirdman, G**., Ghaidan, H., Hyllén, S., Pierre, L., Olm, F., & Lindstedt, S. (2024). Amniotic fluid derived mesenchymal stem cells reduce inflammation and improve lung function following transplantation in a porcine model. The Journal of heart and lung transplantation: the official publication of the International Society for Heart Transplantation, 43(12), 2018-2030. https://doi.org/10.1016/j.healun.2024.08.014

Svobodová, B., Löfdahl, A., Nybom, A., Wigén, J., **Hirdman, G**., Olm, F., Brunnström, H., Lindstedt, S., Westergren-Thorsson, G., & Elowsson, L. (2024). Overlapping Systemic Proteins in COVID-19 and Lung Fibrosis Associated with Tissue Remodeling and Inflammation. Biomedicines, 12(12), Artikel 2893. https://doi.org/10.3390/biomedicines12122893

Bèchet, N. B., Celik, A., Mittendorfer, M., Wang, Q., Huzevka, T., Kjellberg, G., Bodén, E., **Hirdman, G**., Pierre, L., Niroomand, A., Olm, F., McCully, J. D., & Lindstedt, S. (2025). Xenotransplantation of Mitochondria: A Novel Strategy to Alleviate Ischemia-Reperfusion Injury during Ex Vivo Lung Perfusion. The Journal of heart and lung transplantation: the official publication of the International Society for Heart Transplantation, 44(3), 448-459. https://doi.org/10.1016/j.healun.2024.10.033

Hirdman, G., Roijer, A., Paulsson, M., Wierup, Per., Rasmussen, M. (2026). An unusual case of Streptococcus pyogenes infective endocarditis demonstrating the usefulness of repeated echocardiography. IDCases, 44, e02539. https://doi.org/10.1016/j.idcr.2026.e02539.

Abbreviations

ACE2	Angiotensin-converting enzyme 2
ALI	Acute lung injury
AP	Average precision
ARDS	Acute respiratory distress syndrome
AT2	Alveolar type II (cells)
AUC	Area under the curve
BALF	Bronchoalveolar lavage fluid
BLI	Before lung injury
BM	Basement membrane
CLAD	Chronic lung allograft dysfunction
COPD	Chronic obstructive pulmonary disease
DCs	Dendritic cells
DDA	Data-dependent acquisition
DEPs	Differentially expressed proteins
DIA	Data-independent acquisition
EBC	Exhaled breath condensate
EBP	Exhaled breath particles
ECMO	Extracorporeal membrane oxygenation
ECM	Extracellular matrix
EDC	Extended donor criteria
EVLP	Ex vivo lung perfusion
EVs	Extracellular vesicles
FDR	False discovery rate
FEV1	Forced expiratory volume in 1 second
FiO_2	Fraction of inspired oxygen
FVC	Forced vital capacity
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis

ICBs	Immune checkpoint blockers
IF	Immunofluorescence
IPF	Idiopathic pulmonary fibrosis
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LFQ	Label-free quantification
LPS	Lipopolysaccharide
LTx	Lung transplantation
NET	Neutrophil extracellular traps
NK	Natural killer (cells)
NPV	Negative predictive value
NSCLC	Non-small cell lung cancer
OPC	Optical particle counter
ORA	Overrepresentation analysis
PaO_2	Arterial oxygen tension/oxygenation
PEV	Particles per exhaled volume
PFR	Particle flow rate
PGD	Primary graft dysfunction
PPV	Positive predictive value
ROC	Receiver operating characteristic
SSP	Single-sample predictor
VOCs	Volatile organic compounds

Introduction

The Respiratory System

The human respiratory system is a complex structure evolved to manage the constant exchange of gases between the external atmosphere and the internal metabolic environment. This system is highly specialized, balancing efficient transport of oxygen and carbon dioxide with an advanced immunologically active barrier that protects against pathogens, toxins, and particulates. An average adult ventilates over 10,000 liters of air through the respiratory tract daily and facilitates the exchange of more than 400 liters of oxygen (O_2) and carbon dioxide (CO_2) across the alveolar surface[1]. To perform this feat, the lungs consist of approximately 500 million alveoli, providing a total surface area of roughly 70 m^2 [2,3].

Anatomy of the Respiratory Tract

Functionally, the respiratory tract can be divided into two zones: the conductive zone and the respiratory zone. The conductive zone includes organs of the upper respiratory airways, such as the oral and nasal cavities, the pharynx, and the larynx, as well as parts of the lower airways, including the trachea, bronchi, and bronchioles[4]. Starting from the proximal trachea, the airways undergo a branching morphology in which each parent airway splits into two smaller generations, yielding 23 generations according to the Weibel classification described in 1963[2]. Generation 16 in this classification represents the terminal bronchioles, which are the beginning of the distal airways and the end of the conductive airways. The conductive airways serve as the anatomical dead space, transporting air into the respiratory zone but not participating in gas exchange. Instead, it warms, humidifies, and filters the air before it reaches the respiratory zone. The respiratory zone (Generations 17-23) is where gas exchange occurs and includes the respiratory bronchioles (Generations 17-19), alveolar ducts (Generations 20-22), and alveolar sacs (Generation 23)[2].

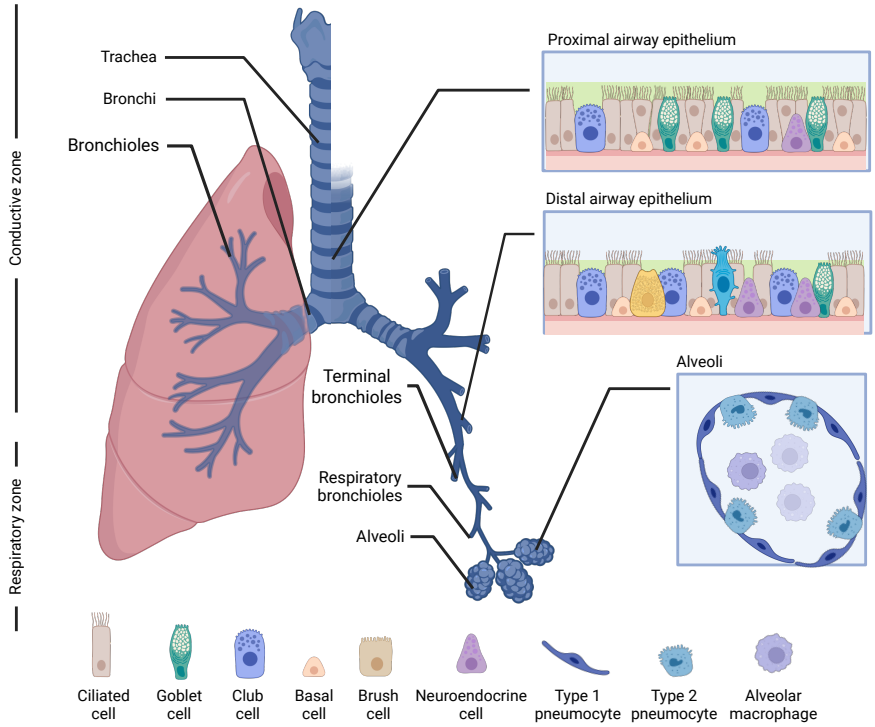


Figure 1: Macroscopic and histological anatomy of the airways.
Functional classification of the lower airways with the conductive and respiratory zone and their subparts. Figure made with biorender.com.

Along this proximal to distal path of the airways, the epithelia of the respiratory tract undergo a transformation from a complex pseudostratified epithelium in the trachea to an extremely thin squamous epithelium in the alveoli. At the trachea and bronchi, the epithelium is composed of a pseudostratified columnar ciliated epithelium with mucus-producing goblet cells, and, as the airways divide, the epithelium progressively decreases in height, into a single-layer simple columnar epithelium in the large bronchioles to a simple cuboidal epithelium in the terminal and respiratory bronchioles[5]. The ciliated pseudostratified columnar epithelium is mainly composed of five distinct cell types: ciliated cells, goblet cells, basal cells, brush cells, and neuroendocrine cells. The ciliated cells are the most abundant and regulate the removal of debris through the mucociliary escalator. Goblet cells are mucin-producing and are replaced by club cells distally along the respiratory tract. Basal cells connect to the basement membrane and function as stem cells for ciliated and goblet cells[6]. Brush cells, meanwhile, are sparsely distributed and are thought to act as chemosensory cells important for innate immune functions[7].

The terminal unit of the respiratory zone, the alveoli, is composed of three primary cell types. The most abundant of these, the alveolar type I pneumocytes, cover approximately 95 to 97% of the alveolar surface area and mainly facilitate the diffusion of gases across the air-blood barrier. They are connected by tight junctions, preventing fluid infiltration into the alveolar epithelium under normal hemostasis. Alveolar type II pneumocytes, meanwhile, are secretory and surfactant-producing cells as well as progenitor cells for the type I and II cells. Lastly, alveolar macrophages are phagocytes that serve as the primary immune defense of the deep lung, traveling freely within the alveoli and septa[6].

Gas Exchange and Pulmonary Ventilation

The blood-air barrier, a thin, single-cell layer only 0.2 μm thick at its thinnest, effectively balances the need for minimal thickness to support gas diffusion with the structural strength required to withstand the mechanical stress of continuous ventilation[8,9]. The barrier consists of three distinct layers. The innermost layer contains type I and type II pneumocytes, the middle layer, composed of a fused basement membrane (BM) of the epithelial BM and the endothelial BM of the capillary wall, and finally, the capillary endothelium, which is composed of a squamous cell lining[10]. The structural integrity of the alveolar barrier is reinforced by the extracellular matrix (ECM) and surfactant within the alveoli, preventing alveolar collapse[11,12].

Gas exchange primarily occurs through passive diffusion of gases along their concentration gradients. This process is primarily explained by Fick's law of Diffusion, which states that the rate of gas transfer ($\dot{V}$) is proportional to the surface area (A), the diffusion constant (D), and the partial pressure difference (P1-P2), and inversely proportional to the membrane thickness (T)[13].

$$\dot{V} \alpha \ \frac{A}{T} \times D \times (P_1 - P_2)$$

The diffusion constant of carbon dioxide is much higher than that of oxygen, allowing it to equilibrate more quickly across the barrier and not require the larger pressure gradient required by oxygen. To maintain gas exchange, ventilation must refresh alveolar air, while perfusion carries away the exchanged gases. For oxygen, equilibration happens within 0.25 seconds, which is about one-third of the time an erythrocyte spends in the capillary. This means that in a healthy person, gas exchange at rest is "perfusion-limited". In conditions such as pulmonary fibrosis and emphysema, on the other hand, which affect thickness and surface area, respectively, gas exchange becomes "diffusion-limited", leading to hypoxia[14].

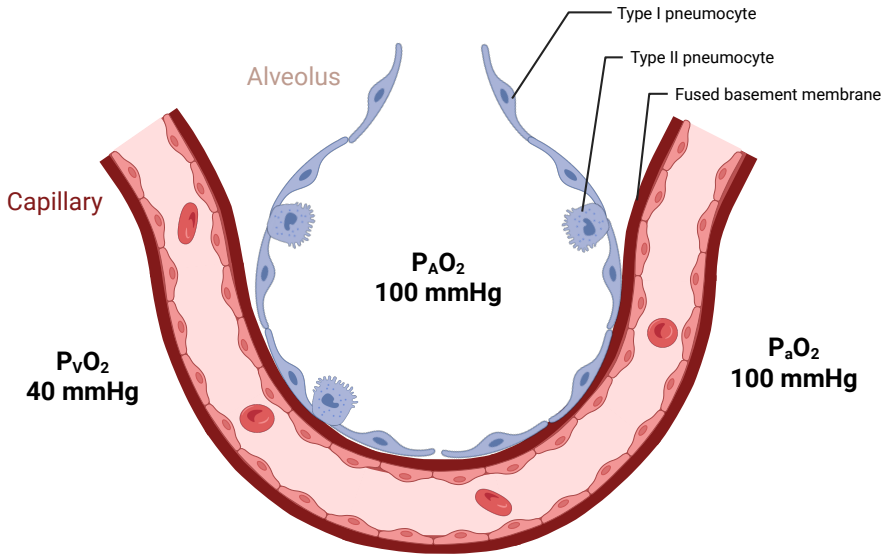


Figure 2: Pulmonary gas exchange in the alveolus.
Oxygen tension around the alveolus. Adapted from Binks, 2022[14] and made with biorender.com.

Normal breathing and thus alveolar ventilation represent an involuntary, rhythmic process regulated by autonomic centers in the brainstem and operate as a negative-pressure pump. Inspiration, and the start of the respiratory cycle, is initiated by contraction of the diaphragm and external intercostal muscles. This causes an increase in intrathoracic volume, a decrease in intrapleural pressure, and, due to the adhesive force of the pleural fluid, the expansion of the lung and decreased intra-alveolar pressure. As a result, a pressure gradient is created between the surrounding atmospheric pressure and intraalveolar pressure, and air rushes into the lungs. Expiration, and the second part of the respiratory cycle, meanwhile, represents a passive process. Relaxation of the diaphragm and intercostal muscles causes the lung tissue to recoil due to the elastic nature of the lung tissue. This increases the intraalveolar pressure relative to atmospheric pressure, causing air to leave the lungs[15]. Mechanical ventilation, an invasive procedure used during and after surgery under general anesthesia and in the intensive care unit (ICU), uses a positive-pressure pump. Air is instead pushed into the lungs during inspiration through an endotracheal tube, while exhalation remains a passive process[16].

The Pulmonary Proteome

The respiratory tract lining fluid (RTLF), also known as the epithelial lining fluid (ELF), forms a thin shield between the pulmonary epithelium and the external

environment. The composition of the RTLF varies markedly throughout the respiratory tract and is mainly composed of mucus in the upper and conductive airways and surfactant from alveolar type 2 (AT2) pneumocytes in the alveolar airways[17]. The composition of this complex molecular protective fluid has been extensively studied using bronchoalveolar lavage fluid (BALF), and the protein pool has been determined to derive from two principal sources: plasma transudation across the alveolar-capillary barrier and local secretions by resident lung cells[18].

Early two-dimensional gel electrophoresis studies in the late 1970s to early 2000s identified fewer than 100 proteins in BALF[19]. With the transition to shotgun liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis, however, the pool has expanded dramatically, with over 1,800 proteins detected in the 2020s[20,21]. Albumin, as in plasma, dominates the BALF proteome and has been suggested as an indicator of barrier permeability[22]. When the alveolar–capillary barrier is disrupted, as in acute respiratory distress syndrome (ARDS), albumin and other high-molecular-weight proteins, including IgM, are commonly measured in BALF as markers of increased alveolar permeability[23,24].

Pulmonary surfactant, in turn composed of approximately 90% lipids and 10% proteins, is another important constituent of the RTLF and contains four surfactant-associated proteins[25]. Surfactant protein A (SP-A) is the most abundant of these and functions as a critical innate immune receptor[26]. SP-B and SP-C are essential for the biophysical property of reducing surface tensions of surfactant[27]. SP-D, similar in structure to SP-A, binds pathogen-associated molecular patterns and enhances phagocytosis by alveolar macrophages. Other important proteins for RTLF function include MUC5B and MUC5AC, large glycoproteins critical for mucociliary clearance and defense against bacterial infections[28].

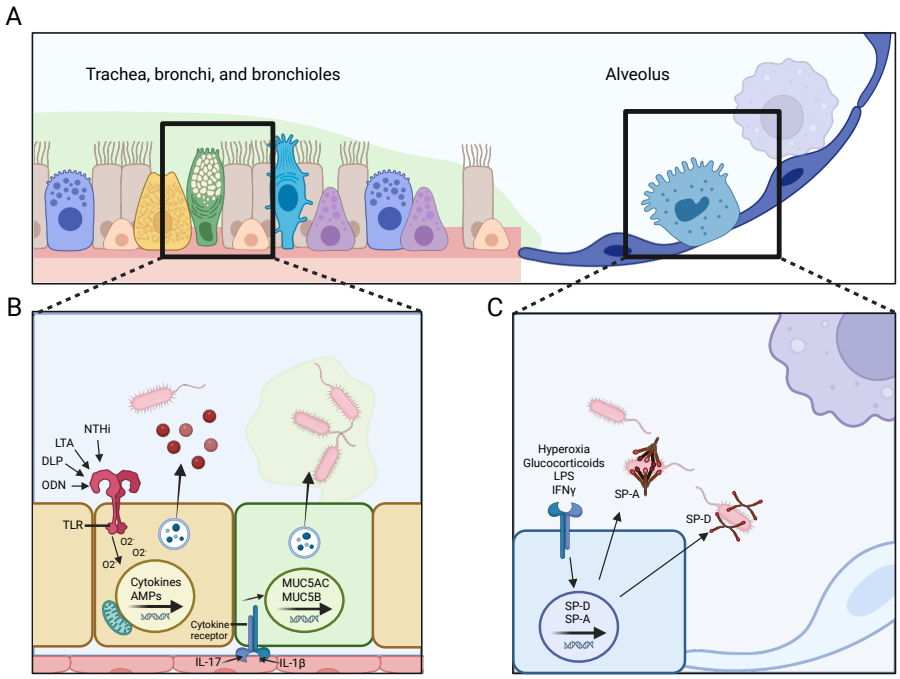


Figure 3: Inducible antimicrobial resistance mechanism of lung epithelial cells. (A) Cells contributing to inducible epithelial airspace defense. (B) Pattern-recognition and cytokine receptors identify local danger signals in the conductive airways, triggering increased barrier and mucociliary functions. This response helps exclude pathogens, boosts the production of antimicrobial peptides and volatile compounds, and promotes the secretion of mediators that recruit and activate leukocytes. (C) Epithelial cells in the alveoli recognize pathogen-associated molecular patterns, detect stress signals, and communicate with lung leukocytes. They respond by modulating barrier function, increasing antimicrobial peptide production, releasing volatile species, and secreting cytokines that activate leukocytes. Adapted with changes and reprinted with permission from Leiva-Juárez et al., © Springer Nature. Figure made with biorender.com.

Another integral part of RTLF defense includes immunoglobulins. Immunoglobulin A (IgA) is the dominant immunoglobulin in airway secretions and is produced locally by plasma cells, where it prevents pathogen adherence and promotes clearance[29]. Beyond adaptive immunity, the lung has been shown to synthesize many complement components. AT2 cells and alveolar macrophages can produce various component proteins that can be activated locally in the alveolar space, producing C3a, C5a, and membrane attack complex (MAC)[30,31]. Additional secreted inflammatory mediators include cytokines, chemokines, and growth factors that help recruit immune cells such as neutrophils, eosinophils, and natural killer cells[32].

Recently, the significance of extracellular vesicles (EVs) signaling in BALF has been recognized. EVs are secreted by various pulmonary cell types into the RTLF, carrying proteins, lipids, and nucleic acids that reflect their cell of origin and the local environment[33,34]. BALF EV studies show that EV proteomes can reflect cancer

status and stage as well as correlate with disease status in various other respiratory diseases, including idiopathic pulmonary fibrosis (IPF), ARDS, and asthma[35–38].

Exhaled Breath as a Diagnostic Source

Historical perspective

Exhaled breath as a source of information dates back to the origins of medicine. Hippocrates, in 460 BC, described *fetor oris* and *fetor hepaticus* as indicative of disease[39]. In 1784, the first breath experiment was performed by Lavoisier and Laplace on a guinea pig (Figure 4). This led to the discovery that mammals consume oxygen and exhale carbon dioxide. A discovery that would lay the foundation of modern biochemistry as well as coin the expression "to be a guinea pig", as Lavoisier later confirmed the finding on himself[40]. Breath research has since been an active area of investigation, with the detection of alcohol in breath in 1874, leading to modern breath analyzers and the urea breath test for Helicobacter pylori testing, among others[40,41].

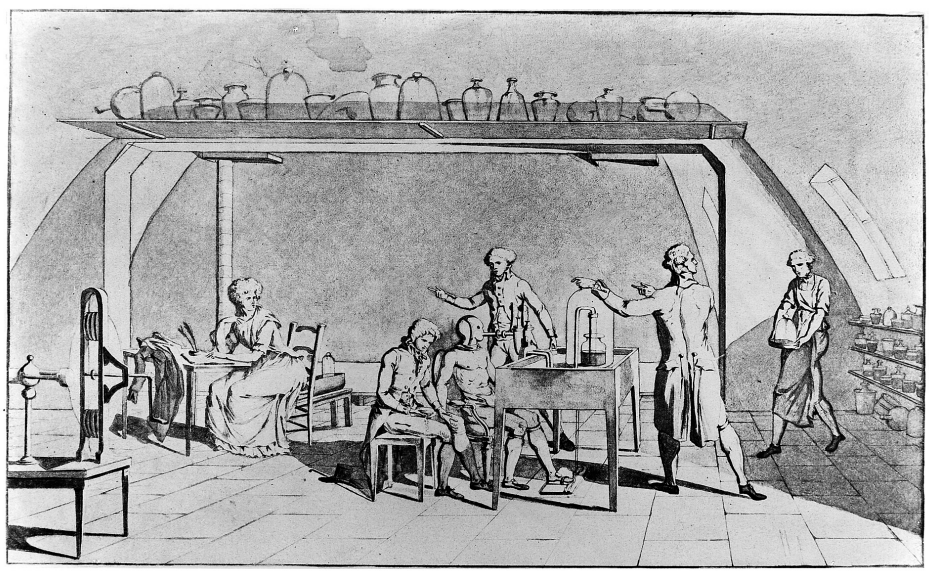


Figure 4: Lavoisier experimenting on human breath in his laboratory. From Wikimedia Commons, licensed under CC BY-SA 3.0.

The field has since developed, and three main types of breath matrices are currently being researched across various health conditions.

Volatile organic compounds (VOCs)

Traditionally, research has been focused on analyzing the volatile gases in exhaled breath. VOCs are produced during various biological processes in the body and discarded through breath. Abnormalities in these biological processes can therefore be detected[42]. VOCs are typically collected in specialized bags, diffusive samplers, or sorbent tubes prior to analysis by gas chromatography mass spectrometry (GC-MS) or gas chromatography ion mobility spectrometry (GC-IMS)[43–49]. These analysis methods, often referred to as electronic noses (E-noses), have been used to identify biomarkers in diseases such as asthma, chronic obstructive pulmonary disease (COPD), lung cancer, as well as hepatic, renal, and gastrointestinal diseases, with mixed but encouraging results[50–55].

Exhaled Breath Condensate (EBC)

In the early 1990s and 2000s, attention began to turn to the non-volatile organic compounds found in exhaled breath[56,57]. Most EBC collection devices collect exhaled breath during normal breathing and cool the air in a cooling container, with the resulting condensation collected as EBC[57]. EBC mostly consists of water (99.9% of total volume) and therefore represents a highly diluted sample[58]. The remaining 0.1% comprises exhaled droplets released from the RTLF by turbulent airflow over the airways[59]. These droplets are thought to arise from the entire respiratory tract, including the larynx, pharynx, and oral cavity, and have not been found to correlate significantly with BALF biomarkers[60]. Research into EBC has been extensive, covering diseases such as asthma, COPD, lung cancer, ARDS, IPF, and cystic fibrosis, among others, utilizing a variety of omics technologies. However, after over 30 years of research, the findings have yielded conflicting results, demonstrated poor reproducibility, and, to date, no clinically viable diagnostic test has been established[57,61].

Exhaled Breath Particles (EBP)

More recently, EBP has been introduced as a novel sampling method from exhaled breath. Briefly, the method involves inhalation of particle-free room air through a two-way valve and exhalation into a thermostat-controlled circuit at 36 °C. In this circuit, an optical particle counter (OPC) continuously draws air samples and counts the particle concentration. Finally, the circuit ends with a two-stage impactor that filters out large particles and collects the finer particles onto a membrane for later analysis (Figure 5). In this way, no dilution of the sample is performed, and sampling can be controlled directly by the total sampled mass, addressing a few of the issues with EBC[37].

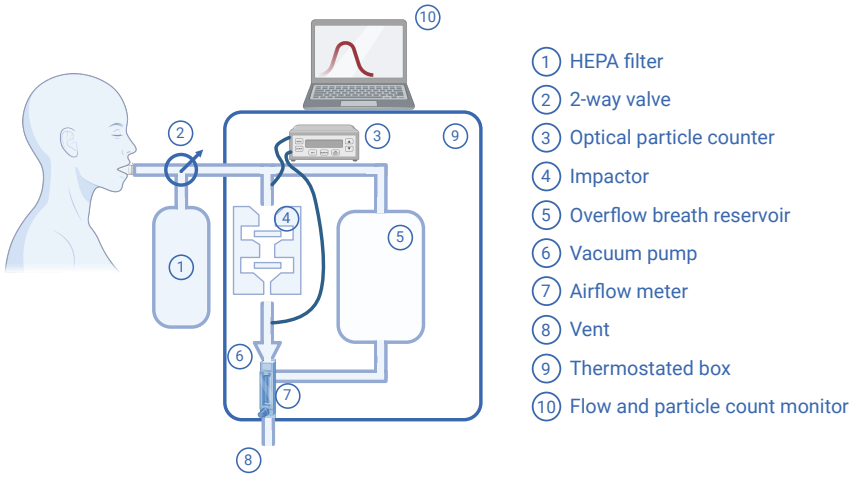


Figure 5: Simplified schematic of the particles in exhaled air (PExA) instrument. The subject breathes through a mouthpiece connected to a two-way, non-rebreathing valve, allowing inhalation of particle-free air through a HEPA filter and directing exhaled air into the instrument. An optical particle counter next samples a portion of the exhaled air. A two-stage inertial impactor sorts particles by size, controlled by a vacuum pump. Any excess exhalation volume is managed by a reservoir and vented out. Figure made with biorender.com.

This approach built on physiological studies suggesting that, apart from turbulent airflow over the RTLF, another important mechanism of particle generation is the closing and subsequent reopening of the distal airways[63,64]. Following this, the airway reopening hypothesis was reinforced by studies demonstrating that deep exhalation elevated particle concentrations, while breath holding diminished particle release and shifted the particle size distribution toward smaller particles, consistent with the gravitational settling of droplets[65]. This was further supported by Almstrand et al., who showed that exhalation to residual volume produced substantially higher particle concentrations than exhalation to closing point or functional residual capacity, indicating that breathing maneuvers that promote airway closure and reopening generate the highest particle concentrations[64]. Subsequent work further suggested that normal tidal breathing and deep exhalations generate particles from different anatomical sites, with the smallest particles arising predominantly from the small airways after airway reopening, intermediate-sized particles originating from the laryngeal region during speech and cough, and large droplets from the upper airways and oral cavity during speech and coughing[66,67]. Thus, by sampling only after a deep exhalation to residual volume, a clean sample from the distal, small airways can be obtained.

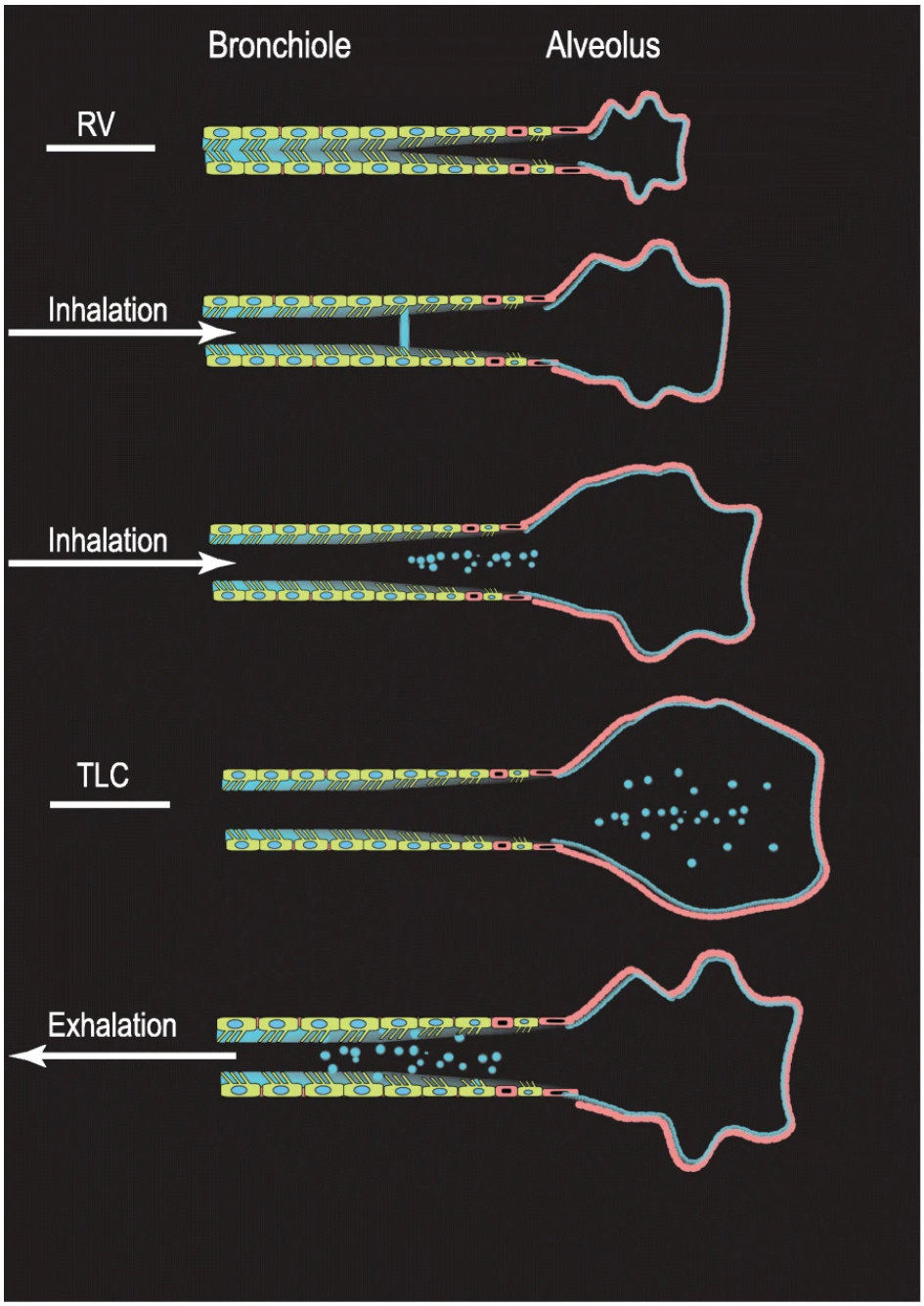


Figure 6: Schematic illustration of the airway reopening hypothesis. Upon deep exhalation to residual volume (RV), the bronchioles collapse into each other, and during inhalation, a thin film of respiratory tract lining fluid forms and then breaks, releasing the resulting particles during exhalation. Reprinted from Bake et al.43 under the Creative Commons CC BY license.

Initial molecular characterization of EBP identified phospholipids and proteins characteristic of pulmonary surfactant, with more than 80% of the identified proteins also detected in BALF studies, supporting the hypothesis of distal airway particle generation[62,68,69]. These include SP-A, SP-B, SP-C, club cell secretory protein (CCSP), various complement factors, annexins, as well as various proteases and protease inhibitors[68]. One notable study compared SP-A and albumin concentrations in EBP with those measured in BALF and bronchial wash (BW), representing distal and proximal respiratory tract lining fluid (RTLF), respectively. A significant correlation was observed between EBP concentrations of SP-A and albumin in BALF, but not in BW, providing additional evidence that EBP originates primarily from the distal airway[70].

To date, EBP has been studied in a few different lung diseases. Early studies have focused on smoking versus non-smoking subjects, finding higher concentrations of EBP and SP-A in current smokers compared to never smokers. This was hypothesized to be due to loss of lung elastic recoil, leading to an increased number of closed and reopened airways, and increased surfactant surface tension, thereby raising droplet concentrations, while increased SP-A levels likely reflect inhalation of toxic smoke[71]. Proteomic studies of smoke-induced changes in EBP highlighted sRAGE, FSTL3, SPOCK2, and protein S as being depleted following long-term and active smoking exposure[72].

Studies on asthma, on the other hand, have found decreased EBP concentrations with increased asthma severity. This is thought to reflect impaired alveolar reopening and small-airway dysfunction. Lower total EBP mass has been associated with small airway hyperresponsiveness, whereas higher EBP mass has been associated with better lung function measures, including forced expiratory volume in 1 second (FEV1), forced vital capacity (FVC), and peak expiratory flow (PEF)[73–75]. On a biochemical level, EBP in asthma has been associated with altered surfactant-related proteins and RTLF makers, particularly lower SP-A levels and depletion of complement and coagulation markers[75–77]. Similar changes are also seen in other chronic airway diseases, such as COPD, where progressive small-airway and parenchymal destruction appear to alter RTLF composition. Similarly to asthma, SP-A levels in EBP samples decrease with COPD severity, and EBP concentrations are significantly lower compared with healthy controls, consistent with surfactant degradation and more proximal airway closure (Figure 7)[78].

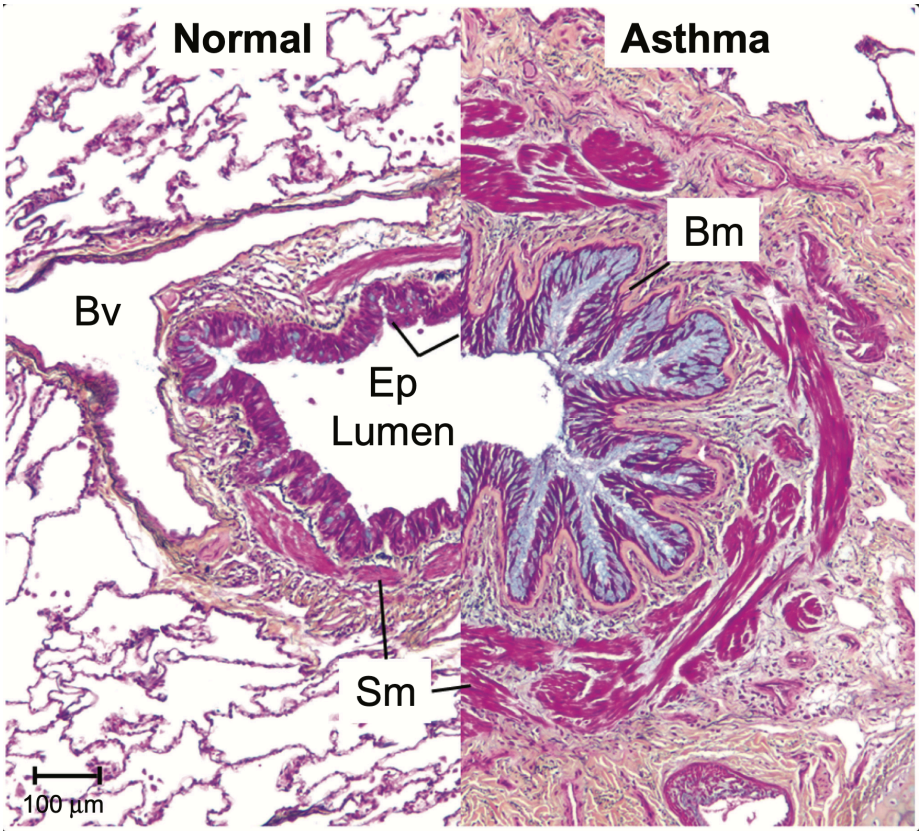


Figure 7. Structural remodeling in asthma. Histology of a healthy lung (left) and a patient with severe asthma (right). The asthmatic lung shows marked narrowing of the airway lumen and thickened epithelium (Ep). Asthma is characterized by airway inflammation and structural remodeling, including thickening of the basement membrane (Bm), smooth muscle hypertrophy (Sm), and increased angiogenesis. Bv, blood vessel. Republished with permission of Dove Medical Press, from Wadsworth et al. 2011[79].

More acute lung pathologies have also been studied. ARDS is a life-threatening, multifactorial syndrome characterized by diffuse alveolar damage, most commonly triggered by pneumonia, aspiration, sepsis, and trauma[80,81]. No specific test currently exists for early identification, and diagnosis is usually based on physical examination, blood oxygen measurements, and radiographic imaging[82]. EBP measurements in mechanically ventilated patients have therefore been proposed as a novel early diagnostic approach. In porcine lipopolysaccharide (LPS) models of ARDS, significant increases in particle flow rate (PFR) were observed before rises in cytokines and other clinical parameters, in some cases by as much as 30 min, and a human LPS inhalation and bronchoscopic instillation model have shown similar increases in EBP together with elevated IL-6 and IL-8 measurable in the EBP[83–85]. This increase is most likely driven by acute inflammatory injury to the distal

airspaces, which disrupts surfactant function and destabilizes small airways and alveoli, leading to increased cyclical reopening during ventilation and consequently greater particle generation.

Lung Diseases and Clinical Challenges

COVID-19

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a respiratory pathogen that causes coronavirus disease 2019 (COVID-19)[86]. It first emerged in Wuhan, Hubei Province, China, in December 2019[87,88]. Following the first genetic sequencing in January 2020, after alarms were raised for an atypical, rapidly spreading pneumonia, it was formally declared a pandemic by the World Health Organization (WHO) in March 2020[89,90]. The respiratory virus causes a broad clinical spectrum of disease, ranging from asymptomatic infection to pneumonia and ARDS with multiorgan dysfunction and death[91]. Disease trajectory is shaped by an interaction of a multitude of factors such as viral characteristics, previous antigenic exposure and vaccination, and host factors such as age, comorbidity, and immunocompromise, amongst others[92].

Structurally, the main proteins of SARS-CoV-2 are the spike, envelope, membrane, and nucleocapsid proteins. Infection of the host starts via viral entry into the host cell through binding of the spike protein to host angiotensin-converting enzyme 2 (ACE2) and priming by the protease TMPRSS2 (Figure 8)[94]. Following entry, the virus undergoes membrane fusion and uncoating, after which the viral RNA is transcribed, structural proteins are synthesized, virions are assembled in the ER-Golgi intermediate compartment, and newly formed particles are released by exocytosis[86].

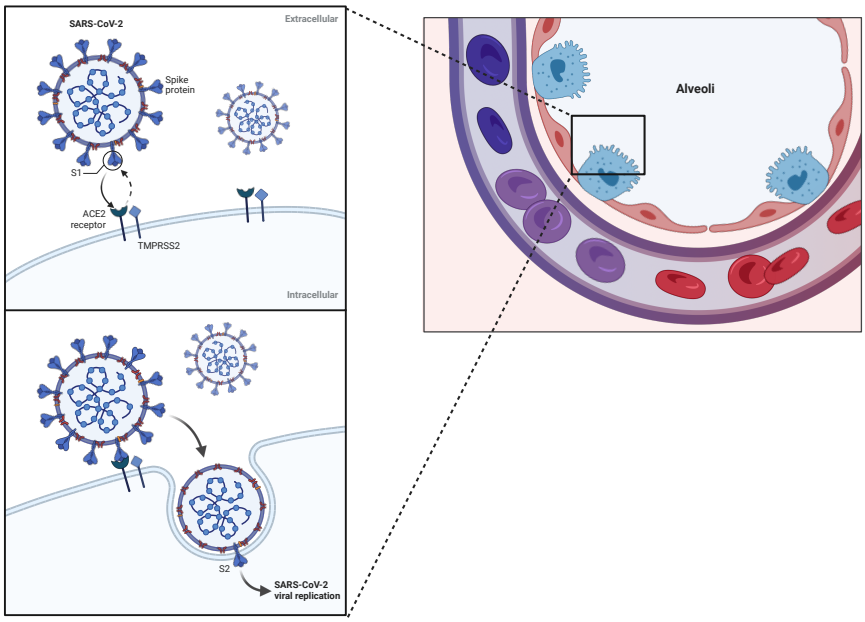


Figure 8: SARS-CoV-2 entry into type II pneumocytes in the alveoli. Viral targeting of type II pneumocytes and entry into host cell through cooperative interaction between spike protein (subunits S1 and S2) and host angiotensin-converting enzyme 2 (ACE2) and transmembrane serine protease 2 (TMPRSS2) protein. Adapted from Ryan et al. 2020 with permission and created with biorender.com

Through its affinity for ACE2 receptors, SARS-CoV-2 primarily infects alveolar epithelial cells, although endothelial cells may also be involved, contributing to cellular injury and the initiation of inflammatory responses[95]. This process disrupts the alveolar-capillary barrier and may permit viral entry into the circulation[96,97]. In a subset of patients, dysregulated antiviral immunity gives rise to a hyperinflammatory state characterized by an exaggerated cytokine response, thereby contributing to severe disease[98]. Furthermore, systemic complement activation may promote microvascular immunothrombosis, further worsening the clinical course in some patients[99,100]. These processes converge at the alveolar-capillary interface and alter the local pulmonary microenvironment, making COVID-19 a highly relevant disease model for studying infection-related biomarkers in alveolar RTLF.

Reverse transcription polymerase chain reaction (RT-PCR) is the most commonly used diagnostic test for COVID-19 and offers high sensitivity and specificity, but it requires access to clinical laboratories[101]. Rapid antigen testing (RAT) is an alternative for at-home testing or in situations where time and resources are limited[102]. Its sensitivity is somewhat lower than that of RT-PCR (70% vs. 90%)[103]. Testing is often performed using nasopharyngeal or nasal swabs[104]. To estimate disease severity, imaging, auscultation, and blood biomarkers are used, but these

provide only indirect information about pathology within the distal lung. Bronchoscopy offers direct access, but it is invasive and unsuitable for repeated sampling[105].

Lung Transplantation

For patients with end-stage lung disease, solid organ lung transplantation (LTx) remains the only treatment option. The field has been evolving ever since Alexis Carrel laid the foundations of anastomosis techniques in 1907 and Vladimir Demikhov performed the first lung transplantation in a dog in 1947. The first human lung transplant, however, was performed by James Hardy in 1963, with the patient surviving for 18 days on an immunosuppressive regimen consisting of prednisone, azathioprine, and cytoablative irradiation of the thymus. Survival after LTx remained abysmal over the following two decades, with the longest recorded survival of 10.5 months in a Belgian patient. This is in unfortunate contrast to a patient who lived for 1.6 years after declining a lung transplantation. It was not until the advent of cyclosporine in the 1980s that the first long-term survivor following LTx was reported, later replaced with tacrolimus, and median survival rose to 4 years. Over the last three decades, however, little progress in long-term outcomes has been made, and survival after LTx remains limited at 6.7 years. Common problems hampering success include primary graft dysfunction (PGD), chronic allograft dysfunction (CLAD), and immunosuppression-related health issues such as cancer and opportunistic infections[106].

LTx is indicated in patients with a rapid decline in pulmonary function and disease progression despite optimal medical care, and the most common underlying diseases include COPD, pulmonary fibrosis, cystic fibrosis, and pulmonary vascular disease[107,108]. Patients who are referred undergo extensive screening for underlying conditions, including psychosocial capacity that hinders long-term survival after LTx[106]. Transplant centers need to carefully balance these factors to allocate the already scarce organs to the patient who might receive the greatest net benefit[109].

There is a worldwide shortage of donor lungs, causing long wait times and waitlist mortality rates of 13.3 deaths per 100 patients[110]. Currently, most donor organs are from donation after brain death (DBD), but donation after circulatory death (DCD) is increasingly being used. Additionally, even when a donor lung is available, it may be unsuitable for transplantation due to injuries like chest trauma, damage from resuscitation efforts, aspiration, ventilator-associated injury, or infection[111]. Traditional selection criteria for donor lungs include age <55 years, arterial oxygenation >300 mmHg (PaO_2), smoking history <20 pack-years, and no aspiration, infection, or a clear chest x-ray[106]. However, as only about 15-20% of potential donor lungs are used, in stark contrast to donor hearts (30%) and kidneys (70%), extended donor criteria (EDC), deviating from one or more criteria, are increasingly used[110]. For some transplant centers, more than half of transplanted

lungs are EDC and show non-inferior outcomes[112,113]. Studies have found that more than 40% of discarded donor lungs could potentially have been suitable for transplantation. Given that the main limiting factor for transplantation is a lack of organs, all efforts at improving donor lung utilization rates are needed[114].

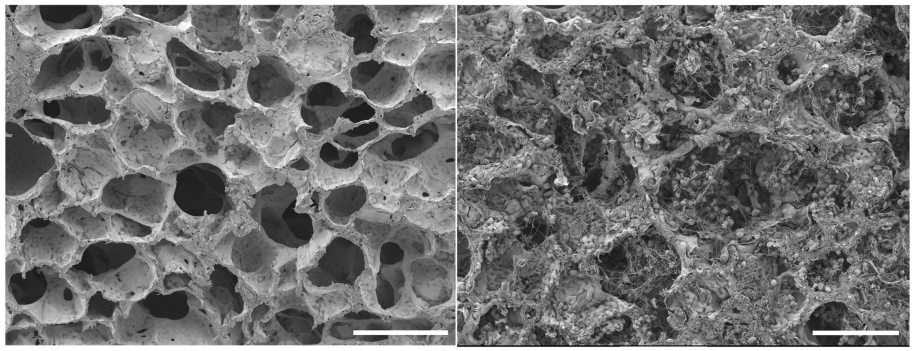


Figure 9: Scanning electron image of alveoli. Healthy alveoli (left) and alveoli damaged by gastric aspiration (right) in a porcine model of gastric aspiration of donor lungs. Scale bars: 100 µm. Image credit Nicholas B. Bèchet, with permission.

Ex vivo lung perfusion (EVLP) has emerged as a transformative platform for increasing utilization rates. Originally developed by Professor Steen and colleagues at Lund University Hospital in 2001, and successfully used for evaluation and subsequent transplantation of initially rejected donor lungs in 2005[115,116]. Briefly, EVLP functions through normothermic perfusion and ventilation of donor lungs outside the body, maintaining aerobic cellular metabolism, enabling functional assessment, allowing for extended warm ischemic time, and serving as a platform to deliver therapeutics to the lung before transplantation[117,118]. However, current selection criteria, including on EVLP, are largely subjective and surgeon-dependent, necessitating more objective lung donor assessment tools and biomarkers[119,120].

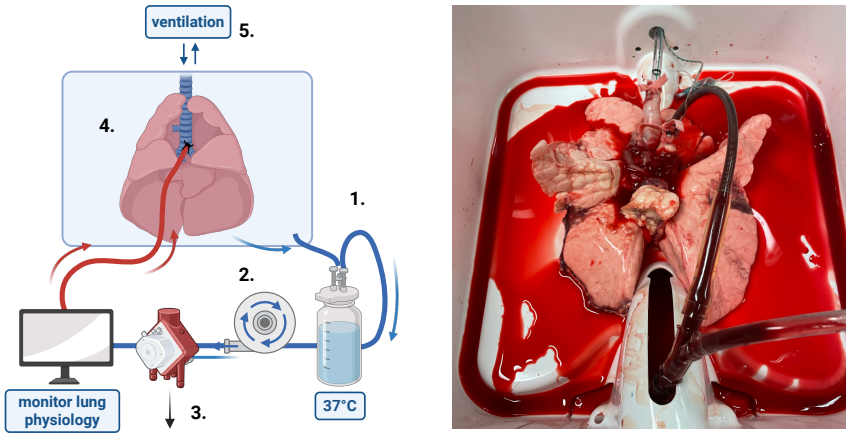


Figure 10: Schematic and image of a typical EVLP circuit in porcine experiments. Blood enters the open reservoir, is pumped through a membrane deoxygenator, heated, and then pumped into the pulmonary artery, where it is oxygenated by mechanical ventilation in the lungs. Figure created with biorender.com, image by the author.

Following transplantation, two main complications account for most of the limited survival rates in LTx recipients: PGD and CLAD[121]. In the acute post-operative setting, PGD is the leading cause of early morbidity and mortality and is largely driven by ischemia-reperfusion injury[122]. It is defined by the onset of pulmonary edema and progressive hypoxemia within the first 72h post-LTx and is graded according to International Society for Heart and Lung Transplantation (ISHLT) criteria[123]. Severe PGD (grade 3) has been reported to occur in 15-25% of recipients and significantly increases ICU-stay and 1-year mortality[122,124]. Molecularly, PGD involves a two-step process, an early stage driven by donor lung macrophages and release of damage-associated patterns (DAMPs) that activate immune signalling via Toll-like receptors, and a later phase mediated by infiltration of recipient neutrophils and lymphocytes[125,126]. Reactive oxygen species generated during reperfusion cause further injury with increased vascular permeability and epithelial cell damage[127]. Treatment remains supportive, with early recognition vital for lung-protective ventilation and, in severe cases, early initiation of extracorporeal membrane oxygenation (ECMO)[128].

Long-term survival in LTx recipients is primarily hampered by the development of CLAD, which occurs in approximately half of recipients within 5 years. CLAD primarily encompasses two phenotypes: bronchiolitis obliterans syndrome (BOS) and restrictive allograft syndrome (RAS)[129]. BOS is characterised by airway fibrosis, and RAS by parenchymal and pleural fibrosis. Pathogenesis includes innate and adaptive alloimmune responses, with early insults, such as severe PGD, viral, bacterial, and fungal infections, and air pollution, driving chronic inflammation. Currently, CLAD is diagnosed based on a persistent decline in FEV1 of more than

20%, and no reversal treatment exists. Lung retransplantation is considered for selected patients with end-stage CLAD but accounts for fewer than 5% of worldwide transplant activity[106].

Lung Cancer

Non-small cell lung cancer (NSCLC) is the predominant subtype of lung cancer, accounting for approximately 85% of all lung cancers, with the rest being small-cell lung cancer. NSCLC is mainly comprised in turn of three different histological subtypes: adenocarcinoma (from mucus gland cells), squamous cell carcinoma (from airway squamous epithelium) and large cell carcinoma (comprised of undifferentiated epithelium-derived tumors)[130]. Lung cancer was the most commonly diagnosed cancer and remains the leading cause of cancer-related death worldwide[131]. Tobacco smoking remains the number one risk factor, but NSCLC in non-smokers is on the rise[132,133]. NSCLC has a broad disease spectrum, from localized curable tumors to metastasized advanced cases with poor survival rates. Survival and disease trajectory is shaped by an interplay of tumor stage, histological subtype, smoking history, comorbidity, and genomic alterations.

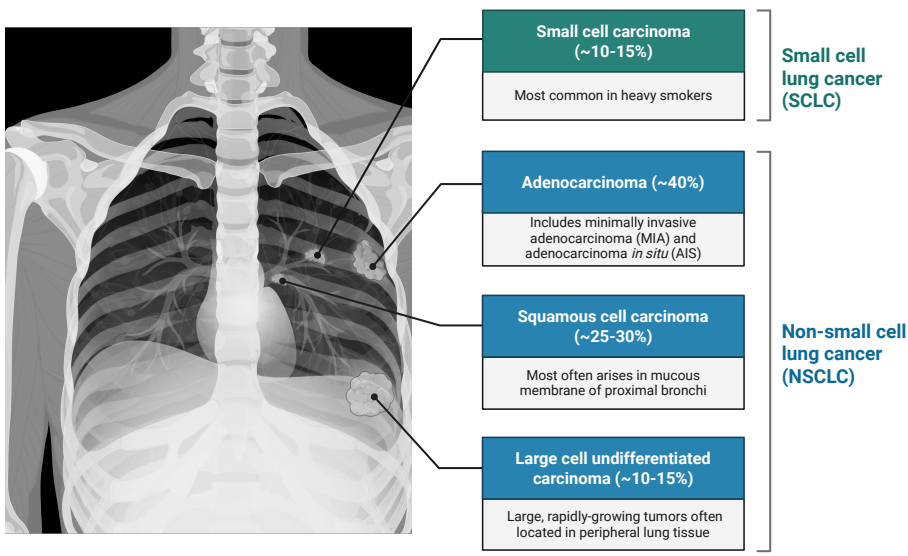


Figure 11: Classification of lung cancer. Lung cancer is divided into two main categories: non-small cell lung cancer and small cell lung cancer. The distinction arose in the 1970s due to the widespread metastatic spread and response to cytotoxic therapies[134]. Figure created with biorender.com

NSCLC develops through progressive accumulation of genomic and epigenetic alterations in the bronchial and bronchioalveolar epithelium. In the distal airways, atypical adenomatous hyperplasia can progress into adenocarcinoma in situ and

invasive adenocarcinoma, while basal cell hyperplasia and squamous dysplasia can progress to squamous carcinoma in situ and invasive squamous cell carcinoma[130]. These processes are carried out through increasing genomic complexity and changes to growth and survival pathways. In adenocarcinoma, therapeutically relevant genomic alterations often involve oncogenic driver genes such as receptor tyrosine kinases EGFR, HER2, and MET, the small GTPase KRAS, or fusion proteins such as ALK, RET, NTRK, and ROS1[135]. Squamous cell carcinoma is less likely to harbor these therapeutically relevant oncogenic pathway activations[130]. Furthermore, the tumor suppressor genes TP53, RB1, STK11, KEAP1, ATM, CDKN2A, and CTNNB1 have been observed to be altered in lung cancer as co-mutations and are thought to be crucial for regulating cancer cell phenotypes, including proliferation, DNA replication, altered cell metabolism, evasion of immune surveillance, and ability to withstand oxidative stress[136–141].

As malignant cell clones expand, interactions with and the importance of the role of the local tumor microenvironment (TME) have been increasingly recognized[142,143]. The lung TME is composed not only of the malignant tumor tissue itself but also of the surrounding vasculature, cancer-associated fibroblasts (CAFs), ECM, and a diverse infiltrate of innate and adaptive immune cells[142]. Stage-dependent immune cell infiltration suggests that the TME actively contributes to cancer progression and may carry prognostic information and therapeutic response[142,144]. Crosstalk between tumor cells and stromal components drives key hallmarks of lung cancer progression. CAFs secrete cytokines that activate pro-tumorigenic JAK2-STAT3 signaling and can suppress T cell function through programmed death ligand 1 (PDL1) and PD-L2[145,146]. The ECM undergoes remodeling through metalloproteinases and lysyl oxidase activity that promotes invasion and acts as a physical barrier to T cell infiltration[147–149] and angiogenesis is primarily mediated through the VEGF pathway to sustain tumor growth and metastatic potential[150].

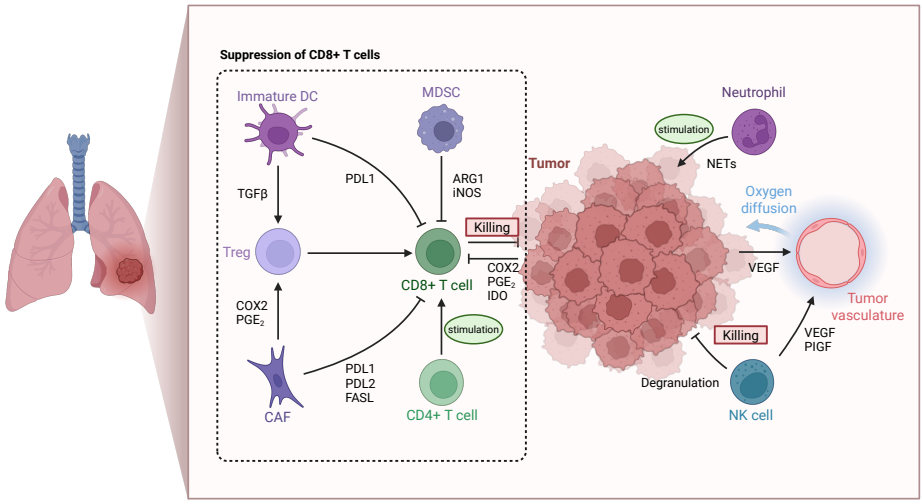


Figure 12: The tumor microenvironment of lung cancer. Suppression of cytotoxic CD8+ T cells is a central hallmark of successful immune evasion in lung cancer and is mediated by a diverse array of cells. Immature dendritic cells (DCs) and cancer-associated fibroblasts (CAFs) from the extracellular matrix stimulate regulatory T cells (Treg) to suppress CD8+ cells. However, DCs and CAFs also directly inhibit CD8+ cells through, among other things, the expression of PDL1 and PDL2. Tumor cells furthermore inhibit CD8+ cells through COX2, PGE2, and PDL1. Neutrophils can stimulate tumor progression through NETs and cytokines. Angiogenesis is promoted by VEGF stimulation from various sources, including natural killer (NK) cells, which in turn exert direct cytotoxic effects via degranulation and NK cell receptors. Figure made with biorender.com and adopted from Altorki et al[142].

The immune landscape of the NSCLC TME is of particular interest. Neutrophils dominate the immune infiltrate and have been found to have a dual, bi-directional role as both anti- and protumor roles. Neutrophils can promote tumor progression through formation of neutrophil extracellular traps, and suppression of anti-tumor immunity[151–153]. Dendritic cells (DCs), myeloid-derived suppressor cells (MDSCs), tumor-associated macrophages (TAMs) and natural killer (NK) cells are all myeloid cell populations of the innate immune system with antitumoral immunity effects but can effectively be reprogrammed by the tumor cells to instead exert immunosuppressive and protumor activity. For example, DCs can be made to produce transforming growth factor-β (TGFβ) causing differentiation of $CD4^+$ T cells into regulatory T cells and in turn suppressing $CD8^+$ antitumor cytotoxic functions[142].

Diagnosis currently relies on clinician suspicion of lung cancer. Diagnosis is made through a combination of laboratory tests, medical history and finally contrast-enhanced computed tomography (CT) scan of the chest while staging is performed through bronchoscopy and invasive biopsies[130]. However, early-stage lung cancer is often asymptomatic and late-stage lung cancer presents with non-specific symptoms such as persistent coughing, dyspnea, hemoptysis and malaise leading to

considerable physician and patient delay. A majority of patients are therefore diagnosed in advanced stages of the disease (Stage III or IV) where 5-year survival rates are below 13%[154]. In contrast, patients diagnosed with localized early-stage disease (Stage I) achieve 5-year survival rates of 73% with surgical intervention. This has led to initiatives to implement screening methods for at-risk populations and large national randomized controlled trials of low-dose (LD)CT scan screening have shown relative reductions in lung cancer mortality of 15-20%[155,156]. Implementation of screening programs has however been limited to a few countries and many questions on harm, risk-benefit, false-positives and costs remain. Non-invasive biomarkers from blood or exhaled breath analysis therefore presents as an optimal rule-in method for improved selection of screening candidates to reduce current drawbacks with large scale LDCT screening programs.

Treatment options include curative surgery or radiotherapy for localized early-stage disease (stage I) and neoadjuvant and perioperative systemic therapy for Stage II-III. Surgical resection is performed either through complete lobectomy or anatomical segmentectomy for patients with significant comorbidities. Systemic therapies include chemotherapy, immune checkpoint blockers (ICBs), and personalized targeted therapies for susceptible mutations. Targeted therapies include Osimertinib, indicated in patients with stage IB-IIIA oncogene-addicted EGFR mutations, and alectinib in ALK-positive tumors. Patients with unresectable stage III disease are treated with chemoradiotherapy and durvalumab, and patients with stage IV disease are treated with systemic therapy based on PDL1 expression levels, contraindications to ICB therapy, and oncogenic driver mutations.

Aims

Study I

To investigate the proteomic composition of exhaled breath particles in patients with COVID-19 and respiratory symptoms, and to evaluate the diagnostic potential of an integrated EBP biomarker panel combined with particle concentration data for non-invasive detection of pulmonary disease.

Study II

To characterize the proteomic profile of exhaled breath particles in porcine and human lung transplant recipients with primary graft dysfunction, validate EBP as a non-invasive alternative to bronchoalveolar lavage, and elucidate the alveolar-capillary barrier changes underlying elevated particle flow rates in PGD.

Study III

To define and compare the molecular and functional responses across three distinct porcine lung injury models, infection, aspiration, and ventilator-induced lung injury, using tissue proteomics, immunofluorescence imaging, and exhaled breath particle analysis, with the aim of identifying injury-specific and shared proteomic signatures relevant to the restoration of damaged donor lungs.

Study IV

To characterize the proteomic alterations in exhaled breath particles and paired tumor tissue from patients with non-small cell lung cancer, establish the biological relationship between the EBP and tumor tissue proteomes, and develop and validate an EBP-based machine-learning classifier for non-invasive detection of NSCLC.

Materials and Methods

Study Designs and Experimental Setup

Study I and IV

Study participants and sample collection

Studies I and IV were observational case-control studies at Skåne University Hospital, Lund. In Study I, patients were recruited from inpatient wards at the infectious disease ward, temporary COVID-19 wards and the emergency department. Patient symptoms, COVID-19 PCR test results and patient characteristics were collected for all participants.

For Study IV, patients with suspected primary NSCLC (clinical TNM stage ≤ cT3N1M0) awaiting surgical resection were recruited, as were age-matched controls of patients awaiting non-emergent surgery for non-pulmonary causes. Exclusion criteria included dementia, severe neurological disease, heart failure (NYHA Class III–IV), ejection fraction below 50%, serum creatinine above 140 µmol/L, and poorly regulated diabetes mellitus. In a subset of the NSCLC patients, tumor and normal adjacent tissue (NAT) were collected intraoperatively, immediately snap-frozen in liquid nitrogen, and stored at -80 °C.

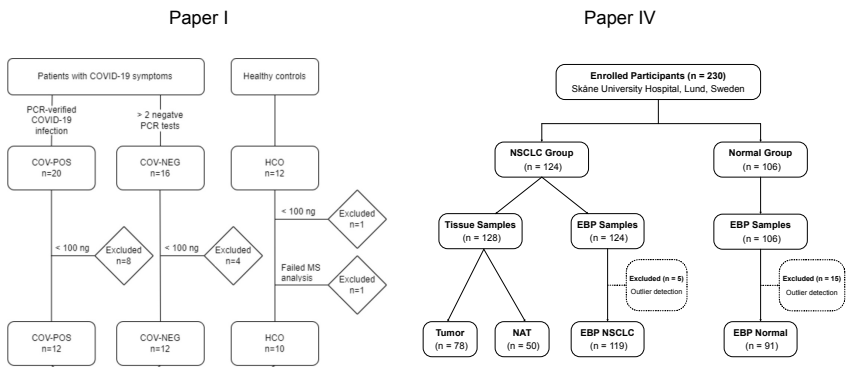


Figure 13: CONSORT diagrams from Study I and Study IV.

Exhaled breath particle collection

In both studies, EBP were collected using a Particles in Exhaled Air (PExA) 1.0 (Study IV) and 2.0 (Study I) device (PExA AB, Gothenburg, Sweden). Participants were instructed to inhale HEPA-filtered air through a 2-way valve for 1 minute, then perform a standardized breathing maneuver consisting of a deep exhalation to residual volume, a 5-second breath hold, and a deep inhalation followed by a final exhalation. Particle number and calculated total mass were continuously recorded by an OPC. In Study I, sampling continued until 120 ng of total particle mass had been collected, and in Study IV, until 200 ng. Particle membranes were stored at -80 °C until LC-MS/MS analysis.

Study II and III

Study design and animal preparation

Studies II and III were experimental porcine studies conducted in accordance with protocols approved by the local Ethics Committee for animal Research.

In Study II, 44 adult farm-raised wild-type American Yorkshire pigs were used in a lung transplantation model. Donor and recipient pairs were matched by blood type and weight, with 22 donors and 22 recipients. Donors underwent pulmonary harvest via median sternotomy during antegrade perfusion with cold Perfadex PLUS solution, and recipients received a left lung transplant. Recipients were immunosuppressed with tacrolimus and methylprednisolone and monitored for 72h. Recipients then underwent a right pneumonectomy of the non-transplanted lung to isolate transplanted lung function. Primary graft dysfunction was staged according to ISHLT guidelines based on PaO_2/FiO_2 ratios and chest imaging.

In Study III, 21 Yorkshire pigs were stratified into three acute lung injury models: 1) intravenous lipopolysaccharide infusion, 2) saline lavage followed by injurious mechanical ventilation, and 3) bronchoscopic instillation of standardized gastric contents. Hemodynamic parameters, including cardiac output by Swan-Ganz catheter and arterial blood gases, were monitored throughout, and ARDS severity was classified according to the Berlin definition. The time points were defined as before lung injury (BLI), 60 min after induction of lung injury (60 min), and endpoint as the establishment of lung injury or death.

Sample collection

Lung tissue biopsies were collected at BLI in Study III and before transplantation in Study II, and at the end of each experiment. Lung tissue samples intended for mass spectrometry were snap-frozen and stored at −80 °C, while samples for histopathological and immunofluorescent analysis were fixed in 10% neutral buffered formalin at 4 °C. In Study II, BALF was collected at the end of the

experiment and stored at -80 °C, and plasma samples were collected at baseline and at the experimental time points.

Exhaled breath particle collection

In both studies, EBP were collected using a customized PExA 2.0 device connected to the expiratory limb of the mechanical ventilator. The device continuously recorded PFR (particles/minute) and collected particles onto a membrane during a total sampling time of 15 min per timepoint. The total accumulated mass at each time point was recorded. Following EBP collection, membranes were excised and stored at -80 °C until further analysis.

Figure 14: Image of a modified PExA 2.0 connected to a mechanical ventilator. Inspiratory air is delivered from the mechanical ventilator to the patient (red arrow), and exhaled air (yellow arrow) passes through the PExA device (red arrow) via the two-way valve (black arrow) to the mechanical ventilator's exhaled air inlet (blue arrow). Reprinted with permission from Broberg et al[157]. © ERJ Open Res.

Experimental methodology

LC-MS/MS

Modern mass spectrometry analysis is a quickly advancing field and has proven to be the most effective method for analyzing complex biological mixtures for their protein content. First developed in 1912 and used for protein measurements in the 1980s, this field has rapidly progressed with more advanced technologies and deeper proteomic insights each year. Traditionally, these measurements are performed on extracted peptides and then matched to the corresponding proteins. A typical workflow consists of three major steps: peptide extraction, peptide separation and MS/MS analysis, and data analysis.

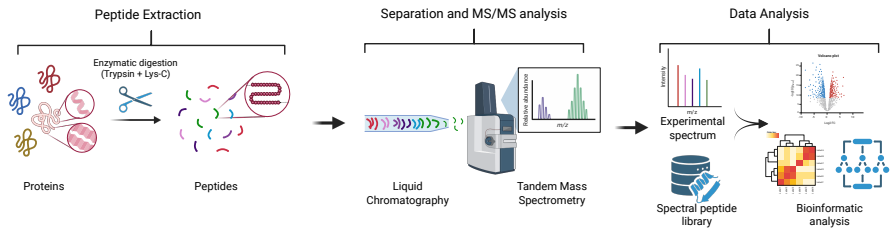


Figure 15: Simplified schematic of a typical LC-MS/MS experiment. Proteins are first digested into peptides using the protease Trypsin, sometimes combined with Lys-C. Next, peptides are separated based on hydrophobicity and passed to the tandem mass spectrometer through electrospray ionization. Experimental spectra obtained from the mass spectrometer are then matched to spectral peptide libraries, and downstream analysis, such as differential expression, clustering, and machine learning classifiers, can be used for biomarker discovery. Figure made with biorender.com

Peptide extraction

Peptide extraction of proteins to be analyzed is usually carried out in a three-step sequence encompassing 1) preparation of the protein for digestion, 2) enzymatic or chemical digestion of proteins into peptides, and 3) cleanup, or desalting and concentration of the resulting solution prior to LC-MS/MS analysis[158].

In studies II-IV, lung tissue was analyzed and required homogenization before peptide extraction. Traditionally, this was done using a pestle and mortar, but more advanced methods have been developed. In studies II-IV, the BulletBlender method, which uses small beads to break up the tissue during centrifugation, was used, creating a tissue lysate.

Following homogenization, the tissue lysate must be solubilized to efficiently release proteins from cellular structures and complexes. In studies II-IV, the extraction buffer contained 2% sodium dodecyl sulfate (SDS) and 50 mM triethylammonium bicarbonate (TEAB). SDS interacts with proteins via ionic and

hydrophobic bonds, altering their secondary and tertiary structures and thereby enabling efficient solubilization of the global proteome[159]. However, SDS is incompatible with LC-MS/MS instruments and analysis and must therefore be removed in subsequent steps. To determine the amount of protein extracted for protease calculations and to standardize protein concentrations, a bicinchoninic acid (BCA) assay was performed. For protein extraction from EBP in studies I-IV, EBP membranes were incubated in a Thermo shaker for 2h at 37 °C and then removed by centrifugation.

Next, proteins must be reduced and alkylated to eliminate disulfide bonds that would otherwise limit protease access. Reduction for tissue and EBP was performed with dithiothreitol (DTT) and alkylation with iodoacetamide (IAA). The lysate was then acidified with phosphoric acid, and SDS was removed with the S-TRAP method. The S-TRAP method enables efficient SDS removal and on-filter protein digestion[160]. Briefly, the S-TRAP method works by capturing the protein particulate suspension created by acidification and the addition of a buffer solution (90% methanol, 100 mM TEAB). The captured protein suspension onto a column filter is washed with multiple steps of the buffer solution.

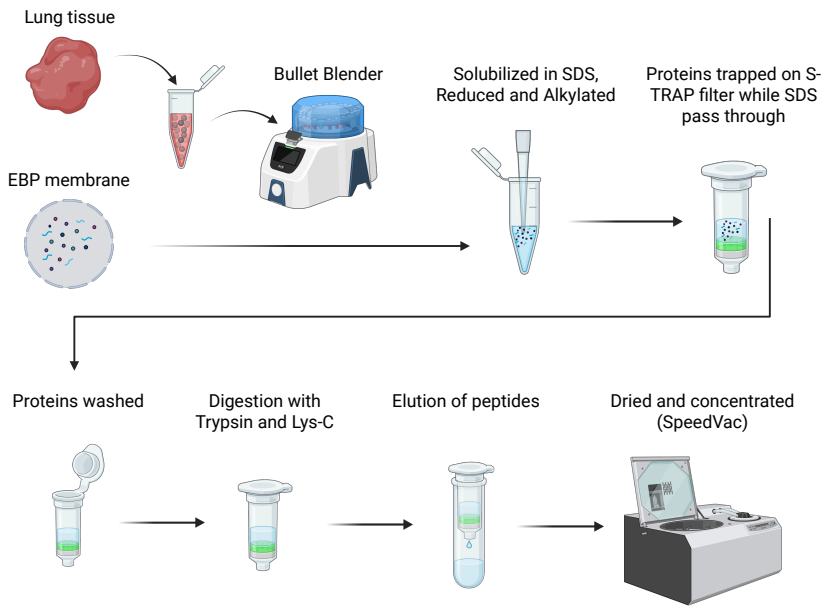


Figure 16: Overview of tissue and exhaled breath particles (EBP) extraction protocol used throughout the thesis. Lung tissue is first homogenized mechanically by a Bullet Blender before solubilization in SDS. EBP are directly solubilized in SDS, and membranes removed. Next, proteins are reduced and alkylated to facilitate enzymatic digestion, and the lysate is trapped on S-TRAP columns with SDS washed off. Digestion of proteins into peptides is done with Trypsin and Lys-C, and the resulting peptides are eluted and finally dried for later analysis by vacuum concentration (SpeedVac). Figure made with biorender.com.

Enzymatic digestion was then performed on-column using trypsin and endoproteinase Lys-C at a 1:50 enzyme-to-substrate ratio each, incubated overnight at 37 °C. Trypsin is the most extensively used protease in bottom-up proteomics and cleaves proteins at the C-terminal side of lysine and arginine residues, and combining with Lys-C has been shown to increase the yield of fully cleaved peptides[161].

Lastly, peptides were eluted from the S-TRAP columns in three steps using buffers with increasing hydrophobicity. First with 50mM TEAB, then with 0.2% formic acid, and finally 50% acetonitrile containing 0.2% formic acid. The combined eluate was then dried by vacuum centrifugation (SpeedVac) and stored for subsequent LC-MS/MS analysis.

Peptide separation and MS/MS analysis

Following peptide extraction, the peptide mixture is then separated prior to mass spectrometric analysis. This step is necessary to deconvolute the highly multiplexed mixture and uses differences in peptide hydrophobicity to separate them on a C18 stationary phase using an increasing organic solvent gradient. The separated peptides are then eluted into the mass spectrometer via electrospray ionization (ESI), where peptides are charged into ions in the gas phase, referred to as precursors[162].

In studies I-IV, separation was performed using the Evosep One LC system. Briefly, dried digested peptides were resuspended in 20 µL of 2% ACN in 0.1% TFA and loaded onto Evotip Pure (Evosep Biosystems, Odense, Denmark). Separation was performed using the 30-samples-per-day (SPD) method.

In studies I-IV, the Evosep LC system was coupled to an MS/MS instrument from the Bruker timsTOF platform. All instruments in this family share a common architecture based on trapped ion mobility spectrometry (TIMS) coupled to a quadrupole time-of-flight (QTOF) mass analyzer. This setup allows for "4D-Proteomics," with analysis of classical mass-to-charge ratio (m/z), chromatographic retention time, and the addition of collisional cross-section (CCS) values, which reflect the size and shape of a peptide in the gas phase.

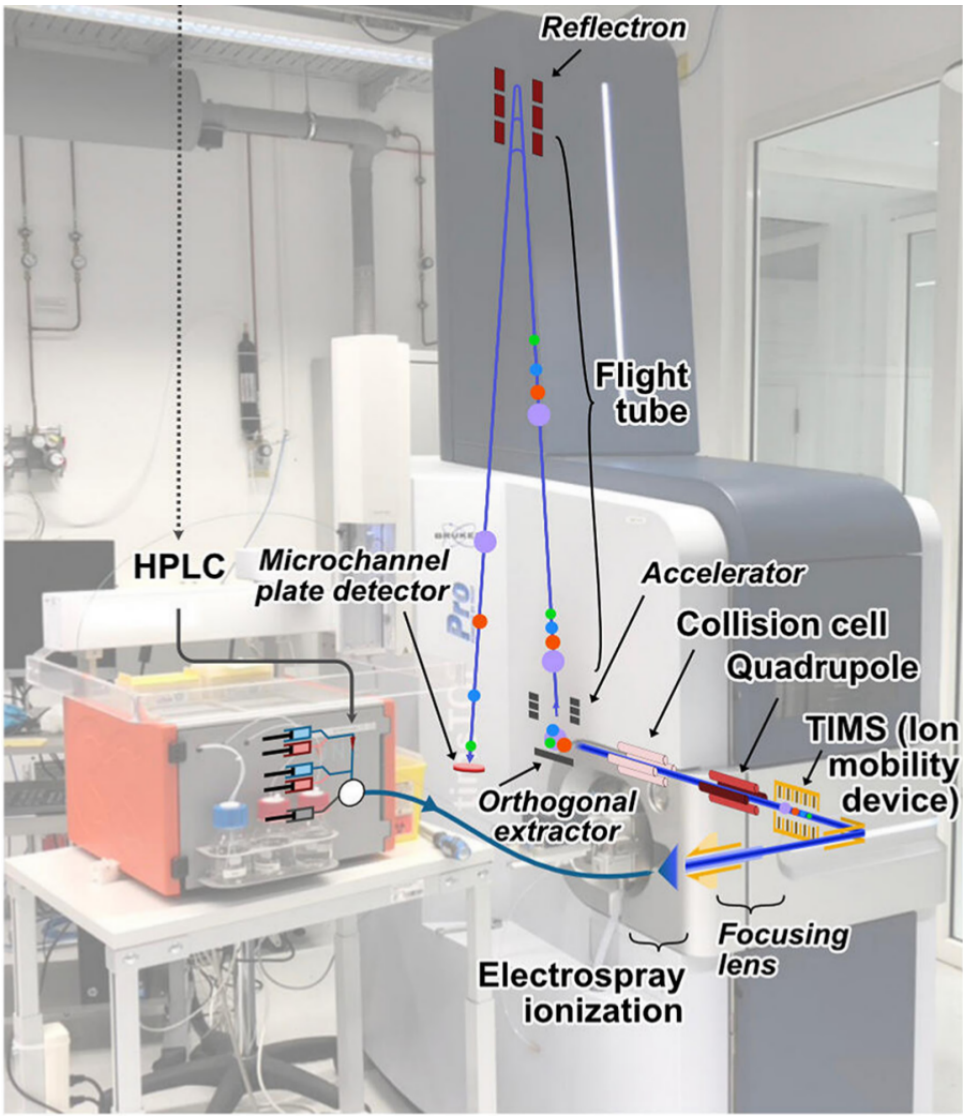


Figure 17: Schematic of a high-performance liquid chromatography (HPLC) tandem mass spectrometry (MS/MS) setup. Peptides are first separated by HPLC and enter the MS/MS through electrospray ionization in the gas-phase. Ions are next pushed through the Trapped Ion mobility device (TIMS) and separated according to their collisional cross-section. Ions are then filtered by a quadrupole mass analyzer and fragmented in a collision cell. The resulting fragment ions are pulsed into a time-of-flight (TOF) analyzer, where they are separated by mass-to-charge ratio (m/z) before reaching the detector. Figure reprinted with permission from Sinha et al.[163] © The Authors and Published by Portland Press Limited.

In studies I-II, the Bruker timsTOF Pro 2 instrument was used for EBP analysis (Study I) and tissue and BALF analysis (Study II). Subsequent studies (studies II-IV) utilized the highly sensitive Bruker timsTOF SCP instrument for EBP proteomics and the Bruker timsTOF HT instrument for tissue-based proteomics.

MS/MS spectra can be acquired in mainly two different modes. In data-dependent acquisition (DDA), the instrument first performs a survey scan (MS1 spectra) and selects the most abundant precursor for fragmentation and generation of an MS2 spectrum. DDA, however, suffers from non-random precursor selection and may miss low-abundance precursors, leading to incomplete data and less reproducible proteome coverage. In data-independent acquisition (DIA), the instrument systematically fragments all precursors within preset fragmentation windows. This generates highly comprehensive and reproducible fragmentations but yields much more complex spectra to analyze and has, over the years, become the preferred method for shotgun analysis of peptides. In Study I, I utilized label-free quantification (LFQ) DDA on EBP, whereas subsequent studies used LFQ-DIA.

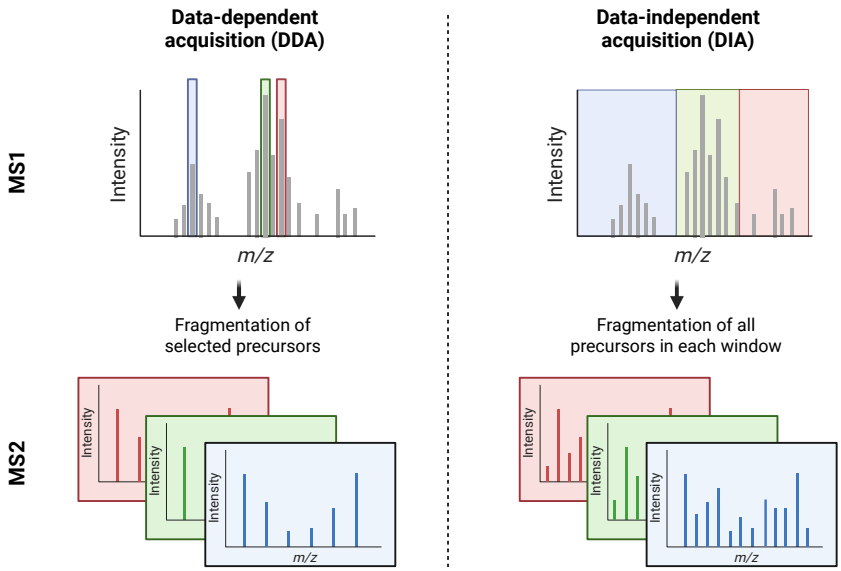


Figure 18: Schematic overview of DDA-MS and DIA-MS. In data-dependent acquisition (DDA), the most abundant precursor ions are selected from the MS1 spectra and fragmented in the MS2 scan. In data-independent acquisition, all precursors within pre-selected scanning windows in the MS1 scan are fragmented in the MS2 scan, yielding more complete but complex data. Adapted from Krasny et al.[165] with permission. Figure created in biorender.com.

Bioinformatic Analyses of LC-MS/MS data

Spectral matching and protein identification

In Study I, raw DDA spectra were analyzed using MaxQuant (v2.0.20) with the integrated Andromeda search engine[166]. Spectra were searched against the UniProt reviewed human proteome and, separately, the UniProt SARS-CoV-2 proteome. In studies II-IV, raw DIA data were processed using DIA-NN (v1.8.1) in library-free mode[164]. DIA-NN utilizes deep neural networks trained on fragmentation and retention time data to generate in silico spectral libraries for EBP data and DDA-generated spectral libraries for tissue data in studies II-III. Database searches were performed against the UniProt human proteome (UP000005640_9606) in studies II and IV, and the UniProt porcine proteome in studies II and III, and supplemented with a common contaminant database[167].

Data filtering, normalization, and quality control

Following spectral matching, downstream data processing was required to remove batch effects and technical artifacts, normalize signal intensities across samples, and filter out proteins with unreliable detection. The specific packages, programs, and strategies evolved across the four studies, reflecting advances in the analytical pipeline.

In Study I, proteins classified as decoy hits, contaminants, or only identified by site were removed, as well as proteins identified in less than 45% of samples in at least one experimental group. Normalization was performed using robust linear regression in NormalyzerDE, and protein-level aggregation was performed using the built-in MaxLFQ algorithm. For the machine learning classifier, missing values were imputed in Perseus (v2.05.0) using a width of 0.3 and a down shift of 1.3, simulating protein abundance under the limit of detection.

In studies II and III, DIA-NN's internal RT-dependent normalization and variance-stabilizing normalization (VSN) and mode-between-protein normalization, as part of the MS-DAP R package, were used for normalization of tissue and BALF data. Proteins were filtered based on identification in at least 60-65% of samples per group, depending on the sample type. Protein aggregation was again performed using the MaxLFQ algorithm. For EBP data, a mixed imputation model in the DEP R package or a hurdle model in the MSqRob2 R package was used due to the high proportion of missing proteins[169,170]. Quality control filtering in Study II removed two porcine EBP membranes and two BALF samples based on manual data inspection.

In Study IV, a much more rigorous quality control pipeline was implemented. Outlier detection was performed using the robust Mahalanobis distance (RMD) filter from the pmart R package[171]. Peptides not identified in at least 75% of samples in at least one group were removed. Aggregation from peptides to proteins was

performed using the robustSummary algorithm in QFeatures, which uses robust regression to reduce the influence of outlier peptides on protein abundance estimates[172]. Normalization was performed in two stages, first, cyclic loess normalization at the peptide level using NormalyzerDE for global intensity differences, followed by EigenMS normalization for extraction batch differences introduced by using 12 extraction batches[173].

Differential expression analysis

In Study I, differential expression analysis was conducted in Perseus using a one-way ANOVA with post-hoc Tukey's test for pairwise group comparisons. Significance was defined as false discovery rate (FDR)-corrected-q-values less than 0.05 and s0 of 0.1. In studies II-IV, differential expression analysis was performed in the MS-DAP R package using the DEqMS and MSqRob2 algorithms, depending on the study and data, and the MSqRob2 algorithm with the QFeatures package for Study IV. DEPs were defined as FDR-adjusted q-values below 0.05 and log2-fold-change thresholds inferred by a bootstrapping algorithm[174].

Functional enrichment and pathway analysis

To gain a deeper understanding of the differential expression analysis, functional enrichment analyses were performed to identify pathways, biological processes, and cellular locations with overrepresented changes.

In Study I, protein-protein interaction networks and Reactome pathways enrichment were conducted using STRING (v11.5), and subcellular localization of DEPs was determined using CellWhere (v1.1)[175,176]. In studies II and III, Gene Set Enrichment Analysis (GSEA) was performed using clusterProfiler (v4.0), and PANTHER was used for Gene Ontology (GO) overrepresentation analysis[177,178]. ECM-specific protein annotation was further performed using MatrisomeAnalyseR to classify ECM proteins by function[179].

In Study IV, functional enrichment was performed using clusterProfiler for both GO overrepresentation analysis (ORA) of upregulated and downregulated proteins separately and GSEA with KEGG database gene sets and protein-protein interaction networks with STRING. Additionally, the EBP proteome was cross-referenced with the Vesiclepedia database to assess the proportion of known EV proteins[180].

Clustering and dimensionality reduction

Unsupervised clustering analysis was used across all four studies to assess the overall structure of the proteomic data and to evaluate the ability of protein expression to separate groups. In studies I-III, heatmaps of differentially expressed proteins were generated using the pheatmap R package. LFQ values were Z-scored, and hierarchical clustering was performed using Euclidean distance. In Study IV, heatmaps were instead generated using the ComplexHeatmap R package on Z-score

LFQ values and with Spearman correlation distance (1 − ρ) with average-linkage agglomeration.

Machine learning classification models

In Study I, a random forest classification model was built using the CARET R package (v6.0-93)[181]. Feature selection was performed in Perseus using ANOVA scores and by minimizing missing values. The cohort was first randomly split 60/40 into training and testing sets, and the model was trained with 1,000 trees in a 10-fold cross-validation setup, repeated 100 times. The number of randomly drawn candidate variables per split (mtry) was tuned using receiver operating characteristic (ROC) area under the curve (AUC) and set to 2. Model performance was evaluated on the held-out test set and reported as accuracy, sensitivity, specificity, and AUC-ROC.

In Study IV, a more extensive machine-learning pipeline was developed. Four single-sample predictor (SSP) classifiers were first benchmarked within a nested cross-validation framework (5-fold, 3-repeat outer loop and a 5-fold, 3-repeat inner loop for hyperparameter tuning) using the rsample and CVMS R package. Next, the multiclassPairs RF model was used for final model training, and feature importance was reweighted based on evidence scores for EV lung cancer origin in the Vesiclepedia database. The final model was trained on the full discovery cohort using optimized settings (gene_repetition = 2, rules_altogether = 50, num_trees = 1,500, min_node_size = 1). A fixed decision threshold was derived by maximizing Youden's J statistic on the training predictions. The locked model was then applied to a separately normalized, batch-independent hold-out validation set, and performance was assessed using the yardstick R package.

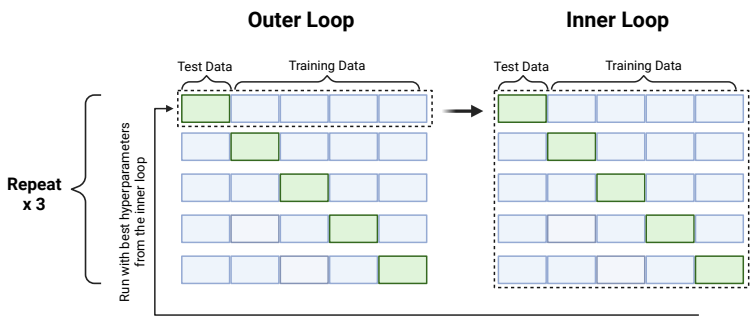


Figure 19: Schematic of the nested 5-fold 3 repeats cross-validation used for final classifier selection. The outer loop (left) evaluates model performance by rotating a held-out test fold (green) across five splits. For each outer split, an inner loop (right) performs a second 5-fold cross-validation on the training data to select the best hyperparameters. Optimal hyperparameters from the inner loop are then used to train and evaluate the model in the outer loop. The entire procedure is repeated three times with different random fold assignments to reduce variance. Figure made with biorender.com.

Histopathology

For histopathological staining (studies II-III), formalin-fixed tissue was processed with ethanol and isopropanol dehydration, followed by paraffin embedding. Thin, 5 µm sections were cut and stained with hematoxylin and eosin and mounted with Pertex (Histolab, Sweden). Brightfield microscopy images were then assessed by multiple experienced, blinded evaluators. Lung injury was scored based on a scale of 0-3 on alveolar wall thickening, hemorrhage, acute inflammation, chronic inflammation, capillary congestion, and alveolar macrophages, and the sum was reported as a lung injury score.

Immunofluorescence staining

For immunostaining analysis (studies II-III), tissue was cut from paraffin blocks with a vibrating microtome. Sections were then deparaffinized and embedded in agarose to provide structural support. Sections were stained with DAPI as a nuclear counterstain, Lycopersicon Esculentum lectin (LEA) to visualize the vascular endothelium, elastin for structural ECM components, and aquaporin-5 (AQP-5) to identify type I alveolar epithelial cells. After incubation and washing, sections were mounted and imaged using a Nikon A1RHD confocal microscopy platform. For quantitative analyses, individual alveoli were captured at random locations across tissue sections from multiple biological replicates to ensure representative sampling.

All immunofluorescent images were analyzed using Fiji software to trace the inner and outer alveolar borders to assess wall thickness, circularity, and morphological quotient (MQ), defined as circularity divided by wall thickness.

Cytokines and chemokines

Plasma samples in Study III were collected at each experimental time point, and plasma cytokine concentrations were measured using a multiplex cytokine kit according to the manufacturer's instructions. A selection of the nine cytokines included in the kit was selected for statistical analysis. Cytokine changes were normalized to each biological replicates BLI value and expressed as $\log_2$ fold change from BLI.

Ethical aspects

All studies included in this thesis were conducted in accordance with the Declaration of Helsinki and were approved by local ethical review boards.

All human participants (Studies I and IV) provided written informed consent, and both studies were approved by the Swedish Ethical Review Authority with Dnr: (Study I: 2018/129; Study IV: 2017/519) and registered at ClinicalTrials.gov (Study I: NCT04503057; Study IV: NCT05395611).

All experimental procedures in Studies II and III were performed according to ethical approval by the Malmö-Lund ethical Committee on Animal Research (Dnr: 5.2.18-4903/16 and Dnr 5.2.18-8927/16 for both Studies II and III).

The animal studies were conducted in compliance with EU Directive 2010/63/EU and adhered to the principles of the 3Rs: Replacement, Reduction, and Refinement, to minimize the use and suffering of animals. As no non-animal alternatives could adequately model the complexity of the lung microenvironment in vivo, animal models were deemed necessary. Sample sizes were kept to the minimum required for statistical validity, and all procedures were refined to minimize pain and distress.

Statistical analysis

The section below excludes all proteomic bioinformatic analyses reported under "LC-MS/MS". Statistical significance for figures was defined as ****$p < 0.0001$, ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, and NS ($p > 0.05$).

Study I

Particle concentration and distribution analyses and patient data statistical analyses were conducted in GraphPad Prism 9 (GraphPad Software, San Diego, CA). For reporting of descriptive statistics, the median and interquartile range (IQR) were used. Kruskal-Wallis test with Dunn's post hoc test was used to compare particle concentrations between groups. For comparison of relative particle size between groups, log-transformed particle bin data were analyzed with a mixed-effects model (REML) with Tukey's multiple comparison post hoc test.

Study II

Descriptive statistics for continuous variables were reported as median and IQR. Statistical differences were tested in GraphPad Prism 9. The Mann-Whitney test was used to compare two groups, and the Kruskal-Wallis test with Dunn's post hoc test was used to compare three or more groups, as the data were not normally distributed.

Study III

All statistical analyses were conducted in GraphPad Prism 10. Statistical differences between cytokine changes in groups and time points were tested with a two-way ANOVA with Tukey's multiple comparisons test. Histology scores were calculated using a one-way ANOVA with Tukey's multiple comparisons, and the change in

PFR from BLI was tested using Friedman's test with multiple comparisons between time points. Correlation analysis between PFR and PaO_2/FiO_2 was performed using Spearman's rank correlation.

Study IV

Descriptive statistics of continuous variables were reported as median with IQR, and categorical variables as n (%). For baseline cohort characteristics, the Wilcoxon rank-sum test was used for continuous variables and Pearson's chi-squared test for categorical variables with FDR correction for multiple testing. For protein intensity correlations, Pearson correlation was performed. All statistical analyses were conducted in R v 4.4.3.

Results

Study I

Patient Demographics and Sampling

In total, 48 participants were recruited for this study in three groups: PCR-verified COVID-19 patients (COV-POS group, n = 20), repeat PCR-negative but COVID-19-symptomatic patients (COV-NEG group, n = 16), and healthy controls (HCO group, n = 12). Median age and sex distributions were similar between the COV-POS and COV-NEG groups: 56 (IQR 53-64) vs 69 (IQR 53-80), whereas the HCO group was generally younger: 44 (IQR 29-46). COV-POS patients had a higher prevalence of obesity (50%) and reported dyspnea more frequently (95%) than the other groups. Symptom profiles were otherwise comparable between the two symptomatic groups regarding fever, throat pain, stomach pain, and myalgia. EBP sampling was performed, on average, on day 7 after a positive COVID-19 test. No adverse events were reported in connection with EBP sampling.

Particle Count Data

Analysis of the optical particle count data revealed significant differences in particle distributions between the groups. COV-POS patients exhaled a median of 11,902 particles per exhaled volume (PEV), and COV-NEG patients a median of 8,159 PEV, compared with a median of just 3,622 PEV in the HCO group. Both groups with respiratory symptoms, therefore, showed significantly increased PEV relative to healthy controls ($p < 0.001$ and $p < 0.01$ for COV-POS and COV-NEG, respectively). Particle size distribution comparisons further showed that symptomatic patients were markedly skewed towards smaller particles. In both the COV-POS and COV-NEG groups, the smallest bin size (median diameter of 0.33 μm) accounted for approximately 33% of total EBP counts in comparison to only 18% in healthy controls. The HCO group also showed a bimodal particle size distribution, whereas symptomatic patients exhibited a right-skewed distribution.

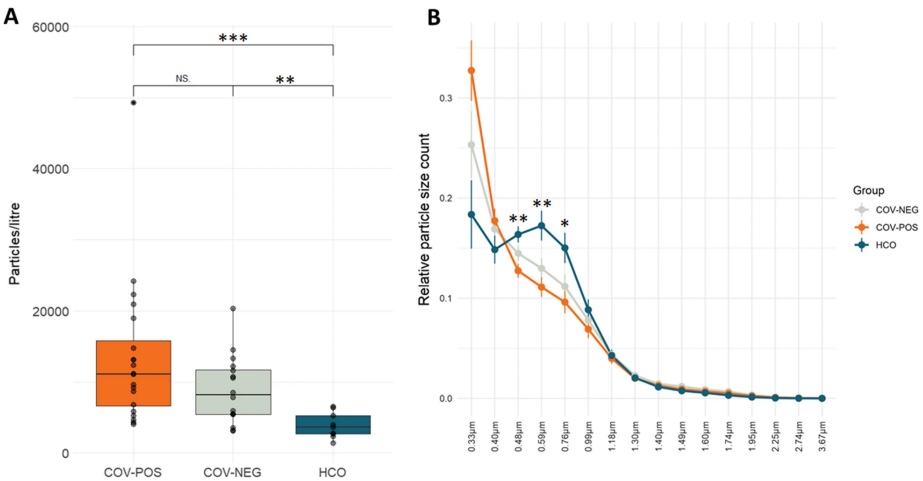


Figure 20: Exhaled breath particle concentrations and particle size distribution between the three study groups. A) COVID-19 positive patients (COV-POS) had a significant increase in particle concentration compared with healthy participants (HCO). B) Particle size distribution was right skewed for COV-POS and symptomatic but COVID-19 negative patients (COV-NEG) as compared with HCO, who presented with a biphasic particle size distribution profile. Reprinted from Hirdman et al.[183] with permission.

Proteomic identification of EBP

From the 48 participants, 34 samples met the threshold set of 100 ng or more of collected EBP mass and were selected for further analysis by LC-MS/MS (12 COV-POS, 12 COV-NEG, and 10 HCO samples). From this, a total of 267 unique proteins were identified across all three groups after excluding known contaminant proteins (mean 110.1, SD 15.8 proteins per sample). Of these, 146 proteins were present in at least 45% of samples of at least one group. Immunoglobulin heavy constant gamma 3 (IGHG3) was the only protein uniquely expressed in the COV-POS group, detected in 50% of all COV-POS samples. No viral SARS-CoV-2 proteins could reliably be detected in any of the samples when searched against known SARS-CoV-2 proteins.

Quantitative analysis of proteins in EBP

Label-free quantification in DDA mode identified 26 DEPs across the three groups. These were found to be predominantly extracellular proteins, but also cell membrane-bound proteins and proteins localized in intracellular compartments. Reactome pathway analysis of the 26 DEPs revealed enrichment in pathways related to the innate immune system and neutrophil and platelet degranulation. Hierarchical clustering analysis of expression levels of the 26 DEPs separated the cohort into three distinct EBP expression profiles. One cluster comprised only COV-POS patients and accounted for 67% of the COV-POS group; the second cluster comprised all three groups; and the third cluster comprised three patients from the COV-NEG group.

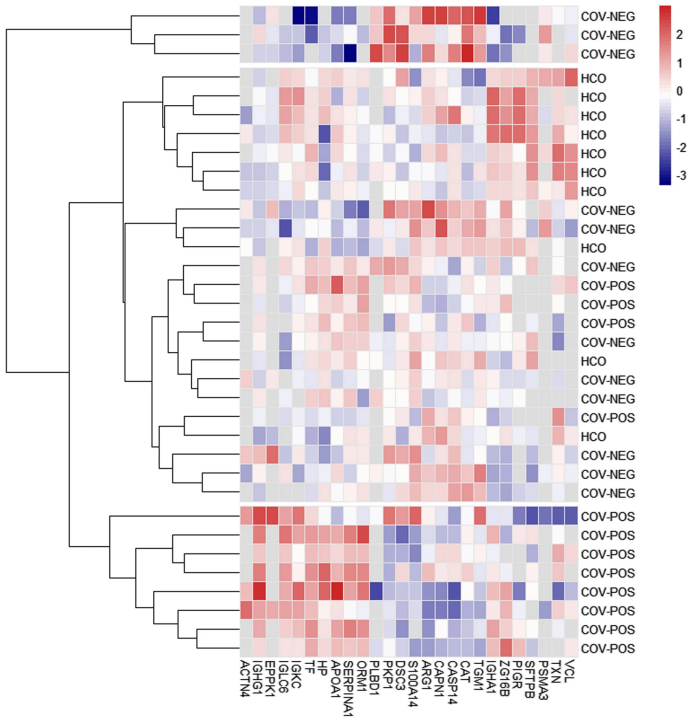


Figure 21: Heatmap of significant protein expression. Samples were clustered into three groups based on protein expression in exhaled breath particles. The bottom cluster consisted only of COVID-19-positive (COV-POS) patients, while the middle cluster consisted of a mixture of COV-POS and symptomatic but COVID-19-negative (COV-NEG) patients, as well as healthy controls (HCO). Reprinted from Hirdman et al.[183] with permission.

Nine proteins were significantly upregulated in COV-POS compared to both COV-NEG and HCO samples. These included three immunoglobulins: IGKC, IGHG1, and IGLC3, as well as epiplakin (EPPK1), serotransferrin (TF), apolipoprotein A-I (APOA1), alpha-1-acid glycoprotein 1 (ORM1), alpha-1-antitrypsin (SERPINA1), haptoglobin (HP), and alpha-actinin-4 (ACTN4). Of these, ORM1, APOA1, and TF were significantly increased in the COV-POS group, while CASP14 and CAPN1 were significantly downregulated. Notably, SFTPB was significantly downregulated in both COV-POS and COV-NEG compared with the HCO group.

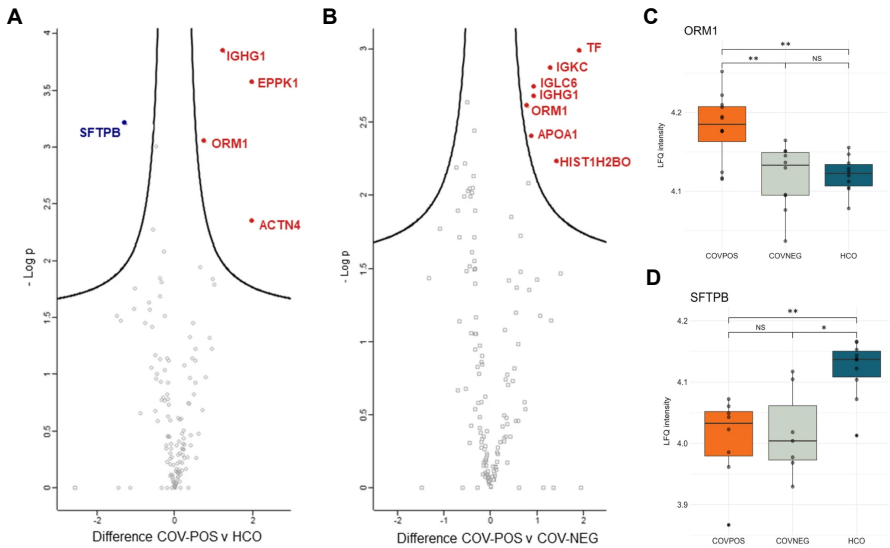


Figure 22: Significantly differentially expressed proteins (DEPs) in exhaled breath particles. A) Volcano plot of DEPs between COVID-19 verified patients (COV-POS) and healthy controls (HCO), B) DEPs between COV-POS and symptomatic but COVID-19 negative patients (COV-NEG). C) Boxplot of label-free quantification (LFQ) intensity of alpha-1-acid glycoprotein 1 (ORM1), an acute phase protein between groups, and D) boxplot of LFQ intensity of pulmonary surfactant-associated protein B (SFTPB), a marker of surfactant health between groups. Adapted from Hirdman et al.[183] with permission.

Machine Learning Biomarker Panel Development

To assess the diagnostic potential of EBP data in acute infectious pulmonary disease, a random forest machine-learning classification model was constructed using 11 DEPs identified across all three groups, combined with each subject's PEV counts. This biomarker panel consisted of the proteins ORM1, IGHG1, CAPN1, CASP14, IGLC6, APOA1, TF, IGKC, EPPK1, SFTPB and IGHA1. The entire cohort was randomly split into training (n = 22) and testing (n = 12) sets, representing a 60/40 split, with subjects classified as either COVID-19 positive (COV-POS) or COVID-19 negative (COV-NEG and HCO subjects). The model was subsequently trained using 10-fold cross-validation, repeated 100 times, with 1,000 trees. In the training set, the ROC-AUC was 0.97 (CI: 0.62-1.00). When applied to the held-out test set the model achieved an ROC-AUC of 0.92 (CI: 0.62-0.99), sensitivity of 75%, and a specificity of 100%.

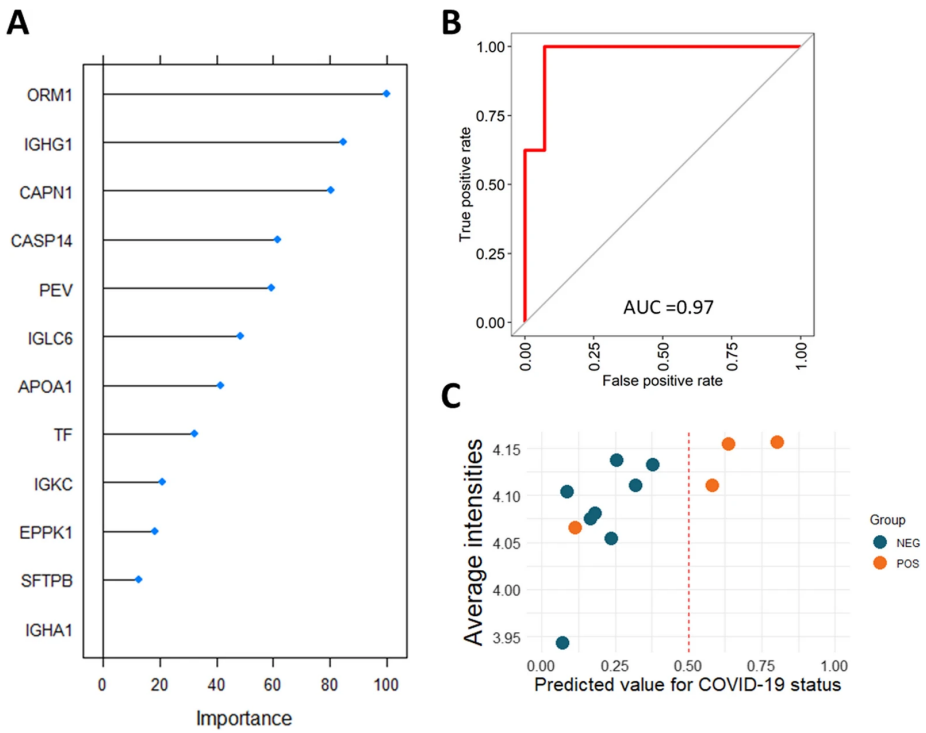


Figure 23: Machine learning model development and validation. A) Variable importance of the biomarker panel. B) Receiver operating characteristics of the training cohort. C) Outcome of the locked model on the validation cohort. Correctly classified samples marked by color and the red line with orange samples (positive) are ideally placed to the right of the red line. Reprinted from Hirdman et al.[183] with permission.

Study II

Increased Particle Flow Rate Correlates with Primary Graft Dysfunction in two models of lung transplantation.

Twenty-two porcine lung transplant recipients were monitored for three days and assessed for PGD on postoperative day 3 according to ISHLT guidelines. Nine recipients developed severe PGD (grade 2-3), and the remaining thirteen did not meet the criteria for PGD (grade 0). PGD scores were further verified with histological analysis, where PGD samples received significantly higher lung injury scores (median 13.00, IQR: 9.75-17.75) compared to the non-PGD group (median 1.50, IQR: 1.00-3.38; $p < 0.0001$), showing signs of damage, including immune cell infiltration, alveolar wall thickening, and capillary congestion in PGD samples.

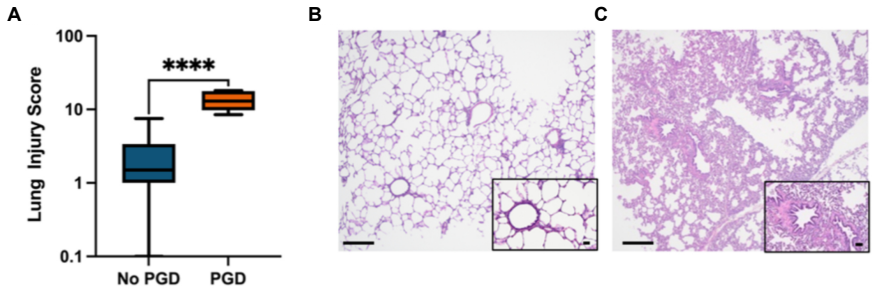


Figure 24: Lung injury scoring and representative histopathology images. A) Blinded scoring of predetermined scoring parameters showed significant differences between porcine with no primary graft dysfunction (PGD) and PGD. B) Representative image of hematoxylin and eosin (H&E) staining of biopsies in transplanted porcine without PGD and C) with PGD. Adapted from Niroomand et al.[184] with permission.

Particle flow rate from the transplanted left lung was next measured using a customized PExA 2.0 device connected to the expiratory limb of the mechanical ventilator. Recipient pigs with PGD (n = 9) exhibited significantly increased PFR of 686.4 (IQR: 449.7-8,824.0) particles per minute (ppm) compared to just 116.6 (IQR: 79.7-307.4) ppm in recipients without PGD (p = 0.0003). Increased PFR in recipients with PGD was next validated in a cohort of 11 human lung transplant recipients, where PGD recipients (n = 5, PGD grade 2) exhibited a significantly increased PFR (461.2, IQR: 284-1,177 ppm) compared with non-PGD human recipients (n = 6; 210.5, IQR: 95-220.2 ppm; p = 0.0424). Furthermore, there was a significant inverse correlation between PFR and the PaO_2/FiO_2 ratio across individual human recipients (Spearman r = -0.7362, p = 0.0128), reinforcing the relationship between disease severity and particle emission from the distal airways.

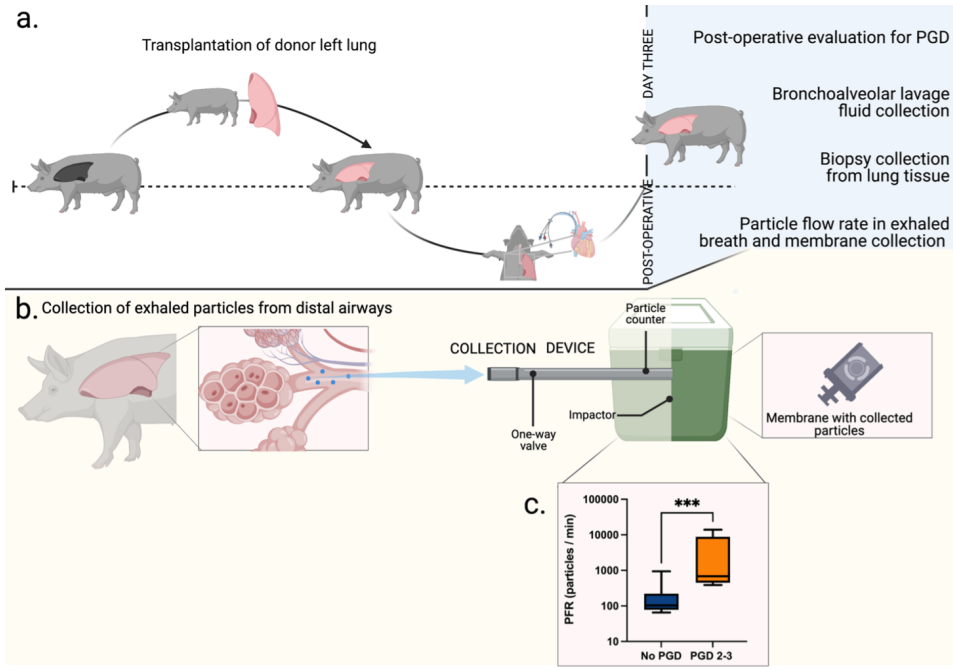


Figure 25: Experimental overview and particle flow rate in porcine recipients. A) Left donor lungs were transplanted, and recipients were monitored for 72h and evaluated for primary graft dysfunction (PGD). Samples including tissue biopsies, bronchoalveolar lavage fluid and B) exhaled breath particles were collected at the end of the experiment. C) Particle concentrations particles/min differed significantly between recipients with no PGD and PGD. Copyright Anna Niroomand with permission. Figure created in biorender.com.

Proteomic Profiling of EBP and Comparison with BALF

We next set out to determine whether the proteins captured by EBP collection reflect the proteomic composition of the distal airways and if changes suggestive of inflammation and leakage could be detected. EBP were analyzed by DIA LC-MS/MS and compared with BALF and tissue proteomes from the same subjects.

In the porcine LTx model, 137 proteins were confidently identified. Among these, seven were significantly overexpressed in the PGD group, and hierarchical clustering of these DEPs indicated that PGD samples tended to cluster together. In the corresponding porcine BALF samples, a considerably more comprehensive proteome was obtained, with 2,418 proteins passing quality filtering; of these, 91 were significantly differentially expressed in PGD, and one was underexpressed. The groups were clearly separated by hierarchical clustering.

In the human EBP samples, 338 proteins were confidently identified, of which 18 were significantly differentially expressed between PGD and non-PGD subjects. In this cohort, a perfect separation between PGD and non-PGD recipients based on the

EBP proteome alone was observed. GO term PANTHER analysis showed significantly enriched terms for acute-phase response and acute inflammatory response proteins (Fold enrichment 22.6 and 16.1, respectively) along with cytoskeletal and filament organization terms.

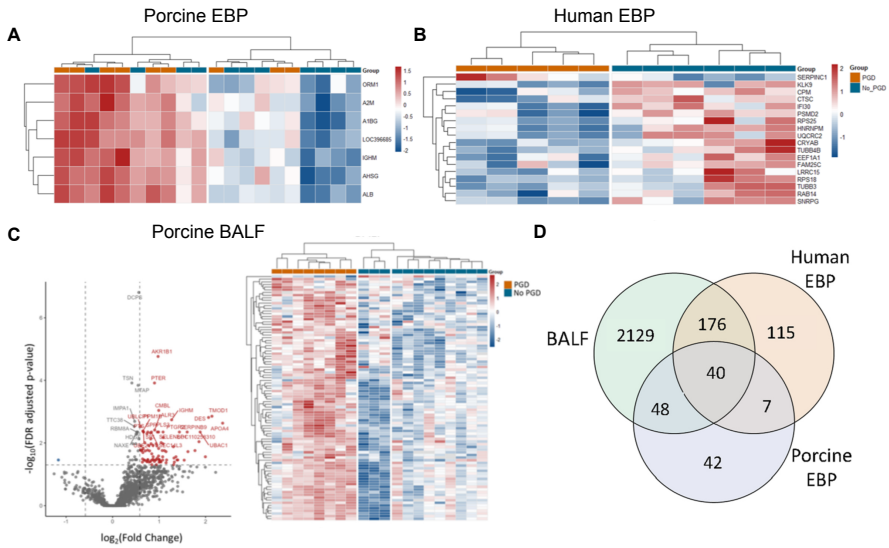


Figure 26: Proteomic expression of porcine and human exhaled breath particles (EBP) and bronchoalveolar lavage fluid (BALF). A) Heatmap of significantly differentially expressed proteins (DEPs) between porcine recipients with primary graft dysfunction (PGD) and no PGD. B) Heatmap of DEPs in humans with PGD and no PGD. C) Volcano plot of DEPs in porcine BALF in recipients with PGD and no PGD. D) Venn diagram showing overlap of protein identifications between BALF, human EBP, and porcine EBP. Adapted from Niroomand et al.[184] with permission.

Furthermore, there was a substantial overlap between the EBP and BALF proteomes. Comparing porcine EBP proteins to porcine BALF revealed an overlap of 88 proteins (64.2% of all porcine EBP proteins), and 216 human EBP proteins overlapped with the porcine BALF, representing a 63.9% overlap, suggesting that EBP in mechanical ventilation partially originates from the same parts of the respiratory tract as BALF.

Tissue Proteomics Reveals Inflammatory, Coagulation, and Extracellular Matrix Alterations in PGD

Next, profiling of the porcine lung tissue proteome in PGD identified 5,206 proteins, of which 302 were significantly differentially overexpressed, and 55 were significantly underexpressed in PGD compared to non-PGD. Hierarchical clustering of these proteins again revealed distinct separation based on PGD status. Gene set

enrichment analysis (GSEA) of the tissue proteome identified enriched biological processes in the PGD group, including regulation of coagulation, hemostasis, blood coagulation, wound healing, defense response, inflammatory response, and immune response. Within the ECM, eleven proteins were significantly enriched in PGD tissue, including metalloproteinase 8 and several known regulators of ECM remodeling. Cross-comparison of the tissue and BALF proteomes with porcine and human EBP datasets revealed not only shared protein identities across sample types but also similar fold-change direction, particularly for proteins grouped under inflammation and immune response, cellular protection and stress response, and ECM organization and cell adhesion.

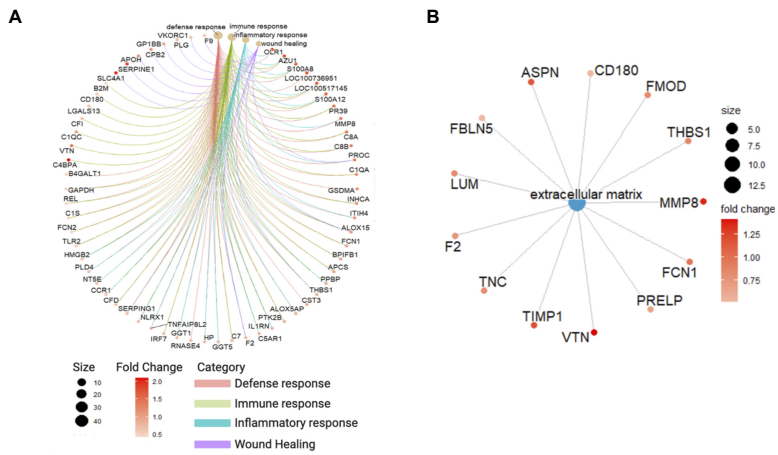


Figure 27: Upregulated biological processes in tissue biopsies from porcine recipients with primary graft dysfunction. A) Gene-concept network plot of the top upregulated pathways in the tissue in primary graft dysfunction (PGD). B) Protein expression for proteins mapped to the gene ontology term “extracellular matrix”. Adapted from Niroomand et al.[184] with permission.

Alveolar-Capillary Barrier Disruption Underlies Elevated Particle Flow Rate

To address the mechanistic question of why PFR is elevated in PGD, the integrity of the alveolar-capillary barrier was next examined at both the histological and proteomic levels.

In porcine lung tissue from PGD recipients, key proteins from both tight and adherens junctions were significantly underexpressed. Among the tight junction components, zonula occludens-1 ($\log_2FC = -0.46$, $q = 0.005$) and occludin ($\log_2FC = -0.83$, $q = 0.04$) were significantly downregulated in PGD, and claudin-18 and junctional adhesion molecule-1 showed decreased, but not-significant trends. In the adherens junctions, junctional plakoglobin ($\log_2FC = -0.42$, $q = 0.005$), catenin-α1

($log_2FC = -0.36$, q = 0.01), and vascular endothelial cadherin ($log_2FC = -0.37$, q = 0.02) were all significantly underexpressed in PGD tissue. Notably, this finding was inverted in the BALF compartment. Adherens junction proteins vinculin ($log_2FC = 0.74$, q = 0.04) and catenin-α1 ($log_2FC = 0.65$, q = 0.03) were significantly overexpressed in PGD BALF, while junctional plakoglobin, PPP2CA, and VE-cadherin all trended upward without reaching significance. This would suggest the shedding or release of junctional proteins from damaged barrier structures into the RTLF.

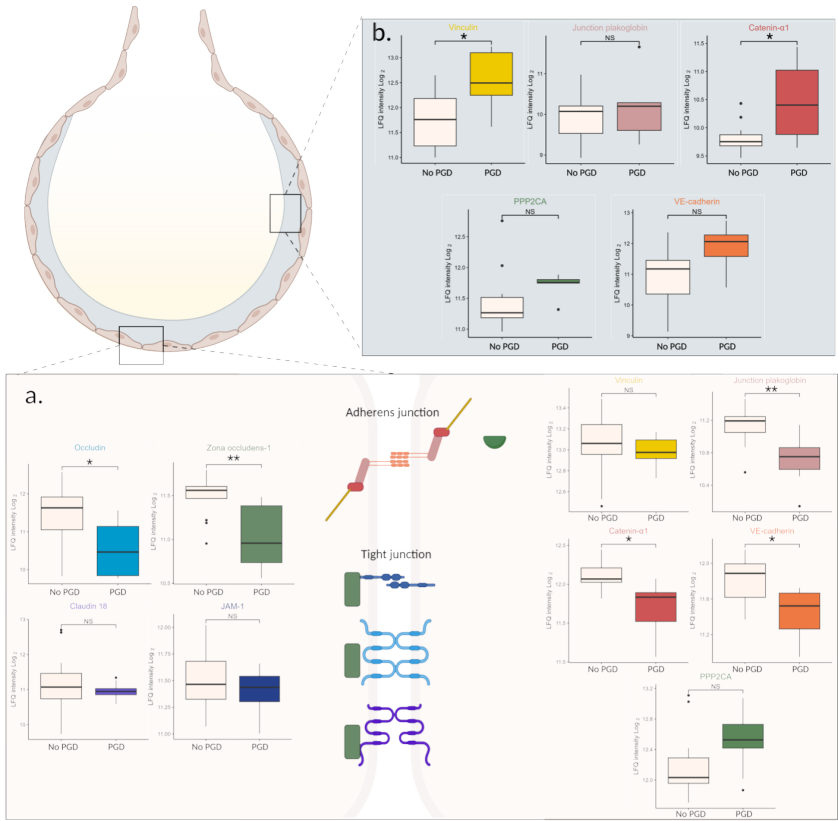


Figure 28: Differential expression of junctional proteins in bronchoalveolar lavage fluid (BALF) and tissue in porcine recipients with primary graft dysfunction (PGD). A) Boxplots of protein expression in tissue with and without PGD in tight junctions (left) and adherens junctions (right). B) Boxplots of protein expression in BALF from porcine with and without PGD. Adapted from Niroomand et al.[184] with permission.

Additionally, evidence of barrier breakdown was found with total protein measurements in BALF, showing a slight increase in the PGD group (2.2 mg/mL, IQR 1.8-10.0) compared with non-PGD BALF (2.0, IQR 1.2-5.9; p = 0.12). More

specifically, IgM, a high molecular-weight serum protein whose presence in the airway is an established marker of alveolar-capillary barrier dysfunction, was significantly overexpressed in PGD BALF fluid. Histological examination further revealed erythrocytes in the airspace of six out of nine PGD recipients, providing direct morphological evidence of barrier breach. Morphological analysis of the alveolar structure reinforced these findings. The alveolar wall thickness was significantly greater in PGD ($p < 0.0001$), while alveolar circularity was significantly decreased ($p = 0.0009$), while the non-PGD group did not differ significantly from pre-transplant samples.

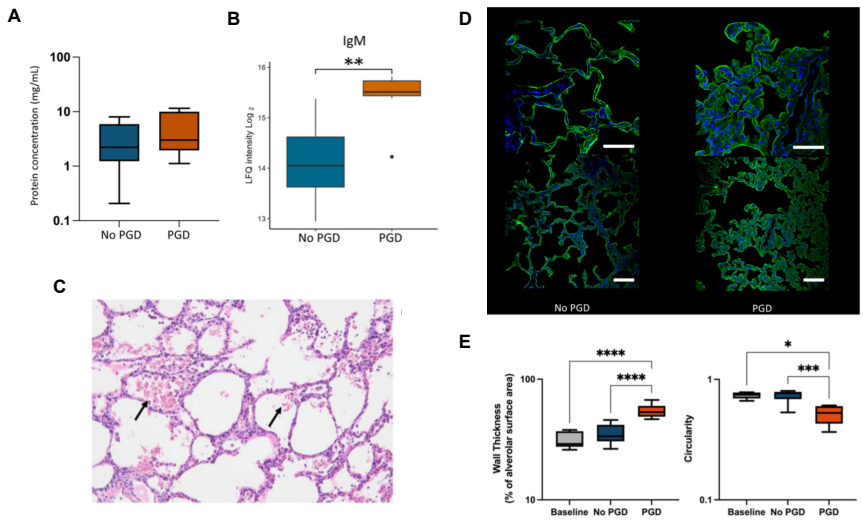


Figure 29: Evidence of barrier breakdown and alveolar changes in primary graft dysfunction (PGD). A) Boxplots of total protein concentrations in bronchoalveolar fluid from porcine recipients with and without PGD. B) Boxplot of immunoglobulin M, a marker of alveolar-capillary barrier leakage, abundance in BALF in PGD and no PGD. C) Hematoxylin and eosin (H&E) stained lung tissue with red blood cells in the airspace (black arrow) in PGD. D) Representative images of immunofluorescence (IF) staining of recipients with no PGD (left) and PGD (right). E) Morphological scoring of IF images showing increased wall thickness and decreased circularity in PGD. Adapted from Niroomand et al.[184] with permission.

Study III

Establishment and Confirmation of Acute Lung Injury Across Three Porcine Models

Three different porcine lung injury models were used to investigate the molecular and functional alterations in donor lung injury. The three models were infection-induced lung injury using intravenous LPS (LPS, n = 7), aspiration-induced injury using bronchoscopically instilled gastric juice (n = 7), and a model of harmful

mechanical ventilation and saline lavage (ventilator-induced lung injury, VILI, n = 7). All three models successfully developed acute lung injury (ALI), as confirmed by arterial blood gas analysis and mechanical ventilation monitoring (decline in pH and PaO_2, dynamic compliance, and increases in $PaCO_2$ and peak inspiratory pressure). Acute respiratory distress syndrome criteria (PaO_2/FiO_2 ratio below 300 mmHg) were met earlier in the gastric group (30 ± 15.0 min), followed by the VILI group (30 ± 75.1 min) and the LPS group (90 ± 42.4 min). Histological scoring confirmed the establishment of lung injury with significant increases in lung injury scores, including immune cell counts, across all groups compared to BLI. Interstitial immune cell infiltration was similar across groups, but fewer immune cells were found in the alveolar airspace in the LPS group.

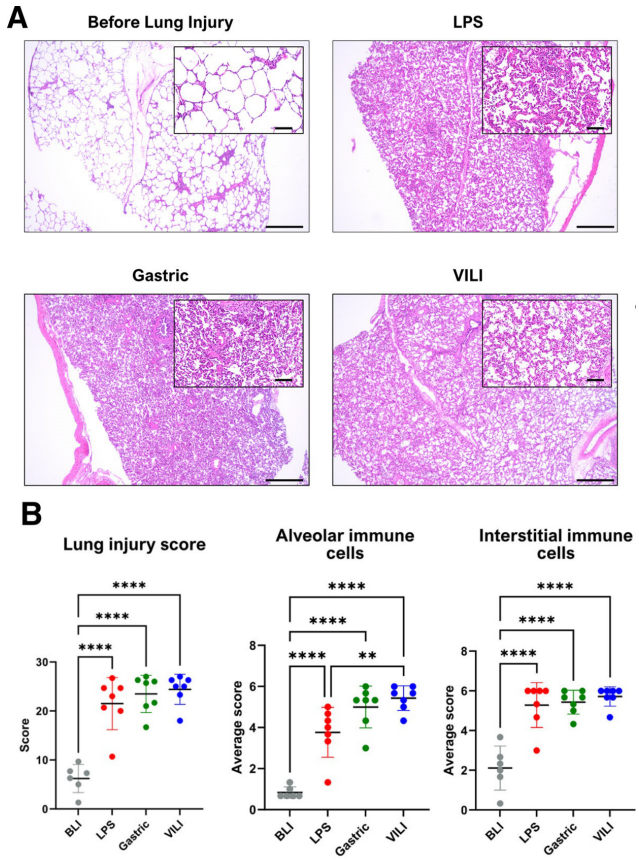


Figure 30: Confirmation of acute lung injury in the three lung injury models. A) Representative hematoxylin and eosin (H&E) stained biopsies of before lung injury and after lung injury in the infection (LPS), gastric aspiration (Gastric), and ventilator-induced lung injury (VILI) groups. B) Blinded lung injury scores of H&E images with significantly increased scores for lung injury score, number of alveolar immune cells, and interstitial immune cells across groups after lung injury. Adapted from Hirdman et al.[185] with permission.

Cytokine profiling in plasma revealed injury-specific inflammatory patterns. The LPS group exhibited the most pronounced systemic cytokine response, with significantly higher levels of tumor necrosis factor-α (TNF-α) and interleukin-10 (IL-10) compared to the other groups. IL-1β was slightly increased in the LPS and gastric groups but remained below the limit of detection in the VILI group, and IL-6 increased significantly compared with BLI values in all groups.

Tissue Proteomics Reveal Injury-specific Pathway Activation and Consistent Extracellular Matrix Alterations

LC-MS/MS analysis of lung tissue identified a mean of 6985 ± 442 proteins per sample. Differential expression analysis between BLI and endpoint samples revealed markedly different numbers of DEPs across groups. 736 DEPs in the gastric cohort, 100 in the VILI cohort, and only 33 in the LPS cohort. Hierarchical clustering analysis of DEPs clearly separated the three injury types and the BLI from lung-injury samples, indicating distinct proteomic profiles in each lung-injury model.

GSEA analysis revealed both common and distinct pathway activations across groups. Defense response, immune system process, cell chemotaxis, leukocyte migration, and leukocyte cell-cell adhesion were increased in all lung injury models. However, injury-specific pathway changes were also observed. The gastric group uniquely showed activation of T-cell proliferation and suppression of complement activation, while the VILI group uniquely showed activation of monocyte chemotaxis-related proteins along with suppression of lipid metabolism and B-cell-mediated immunity. The LPS group uniquely exhibited activation of innate immune and stress response pathways.

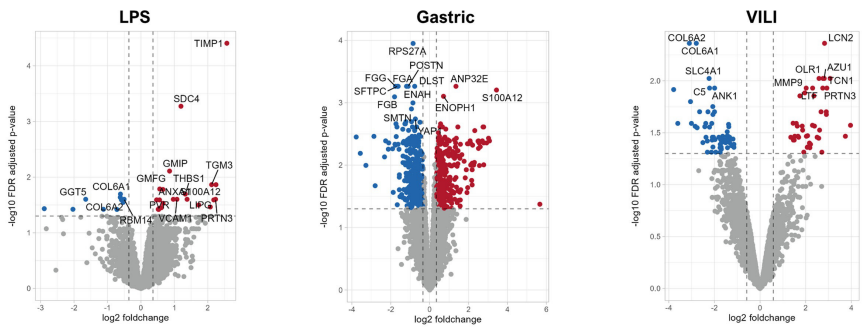


Figure 31: Differential proteomic expression analysis in tissue across three models of lung injury. Shows volcano plots of changes in protein abundance from before lung injury in infection (LPS), gastric aspiration (Gastric) and ventilator-induced lung injury. Adapted from Hirdman et al.[185] with permission.

Across all three models, a disproportionate number of DEPs were ECM-related proteins. This was most pronounced in the LPS group, where ECM-related proteins accounted for 42.4% of the differentially expressed proteome despite constituting only 4.5% of the total identified proteome. Changes to the ECM proteome were characterized by the loss of ECM glycoproteins, collagens, and ECM-affiliated proteins, accompanied by upregulation of ECM regulators. Five ECM-related DEPs were shared across all three lung injury models. These were collagen type VI alpha chains 1, 2, and 6 (COL6A1, COL6A2, COL6A6) as well as the calcium-binding proteins S100A8 and S100A12. This consistent downregulation of multiple COL6 chains, key components of the lung basement membrane involved in tissue maintenance, cell adhesion, and migration, suggests that early COL6 degradation may serve as a universal indicator of the onset of acute lung injury.

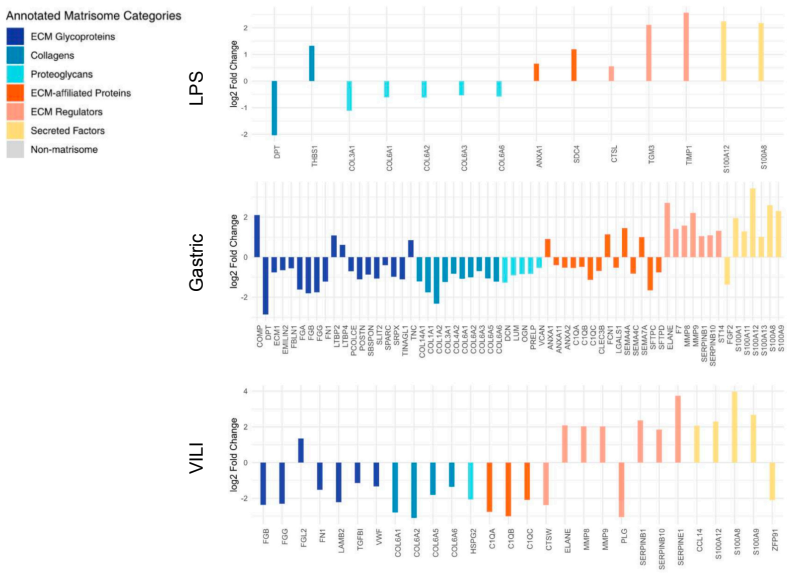


Figure 32: Changes in extracellular matrix (ECM) constituents and regulators in lung injury. Barplots of differentially expressed proteins annotated as belonging to the ECM in three different models of lung injury: infection (LPS), gastric aspiration (Gastric), and ventilator-induced lung injury (VILI). Adapted from Hirdman et al.[185] with permission.

Immunofluorescence Imaging Uncovered Spatially Resolved ECM Remodeling

Multiplex immunofluorescence (IF) imaging of BLI and post-injury lung samples revealed both shared morphological changes and injury-specific structural remodeling. All injury models exhibited disruption of tissue morphology, including increased bronchoalveolar scaffolding coverage, reduced airspace, and uniform thickening of the alveolar walls.

Targeted imaging of elastin expression revealed injury-specific patterns not captured by global proteomic changes. Elastin, an ECM component, was particularly enriched in the gastric and LPS groups with intensified fluorescence and vessel wall thickening around blood vessels. AQP-5, a water channel protein linked to tight junctions present on the apical surface of pneumocytes, showed no significant changes in global proteomic expression. However, IF imaging revealed a shift from continuous, uniform AQP5 distribution in healthy lungs to a patchy, discontinuous pattern in injured lungs. This was seen most significantly in the VILI model.

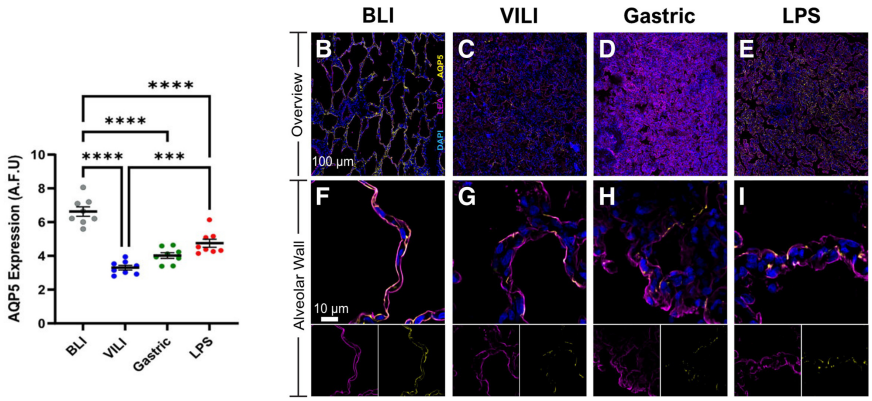


Figure 33: Acute lung injury reduces expression of aquaporin-5 (AQP-5). A) Quantification of AQP-5 expression in the alveoli before lung injury (BLI) and in three models of confirmed lung injury: ventilator-induced lung injury (VILI), gastric aspiration (Gastric), and infection (LPS). B-I) Representative images of the bronchoalveolar scaffold from BLI and lung injury samples stained for AQP-5 (yellow). Adapted from Hirdman et al.[185] with permission.

Particle Flow Rate From the Airways Increased with Progression of Lung Injury

PFR, measured with a PExA 2.0 device connected to the expiratory limb of the mechanical ventilator, increased significantly over the experimental timeline in all three ALI models. To evaluate whether increases in PFR correlate with pulmonary dysfunction, PFR was assessed for correlation with the PaO_2/FiO_2 ratio. A significant negative correlation was found in the LPS group ($r = -0.79$, $p < 0.0001$) and in the gastric group ($r = -0.65$, $p = 0.0037$), indicating that higher PFR rates are associated with progression of lung injury. However, no significant correlation was found in the VILI group ($r = 0.26$, $p = 0.2664$).

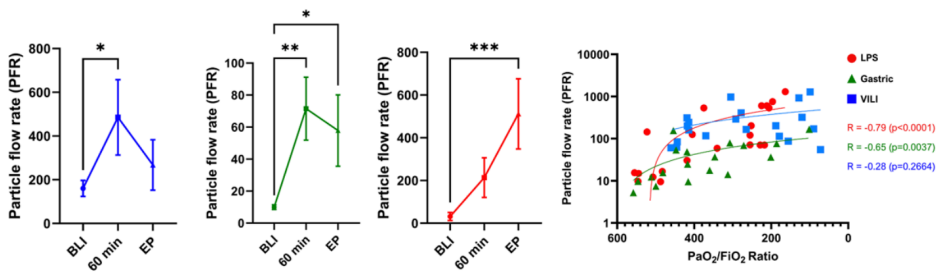


Figure 34: Particle flow rate (PFR) increased significantly during induction of lung injury and correlated with lung function. Shows increased PFR across the three models of lung injury: infection (LPS), gastric aspiration (Gastric), and ventilator-induced lung injury (VILI), and correlation plot with PFR (y-axis) and arterial oxygen partial pressure to fractional inspired oxygen ratio (PaO_2/FiO_2). Adapted from Hirdman et al.[185] with permission.

Proteomic Composition of Exhaled Breath Particles Reflects Injury-specific Molecular Signatures

LC-MS/MS analysis of EBP samples identified an average of 576 ± 135 proteins per sample. Differential expression analysis of BLI and endpoint samples revealed 17 DEPs in the LPS group, 12 in the VILI group, and 106 in the gastric cohort.

Protein-protein interaction analysis of the EBP DEPs revealed strong functional connections among DEPs within each group. GO term enrichment analysis found DEPs were significantly enriched for neutrophil and platelet degranulation pathways, as well as complement and coagulation cascade pathways in the LPS and VILI groups (FDR = 1.42×10^{-7} and 6.53×10^{-6}, respectively). The gastric cohort, meanwhile, showed significant upregulation of cell junction proteins (FDR = 8.08×10^{-7}). Interestingly, most DEPs in the gastric EBP samples were annotated as belonging to the cytoplasm and the alimentary canal. The cytoplasmic origin of these proteins could suggest chemical rupture of cell walls and release of intracellular components into the RTLF, while the alimentary canal annotation likely reflects residual contamination from the aspiration injury.

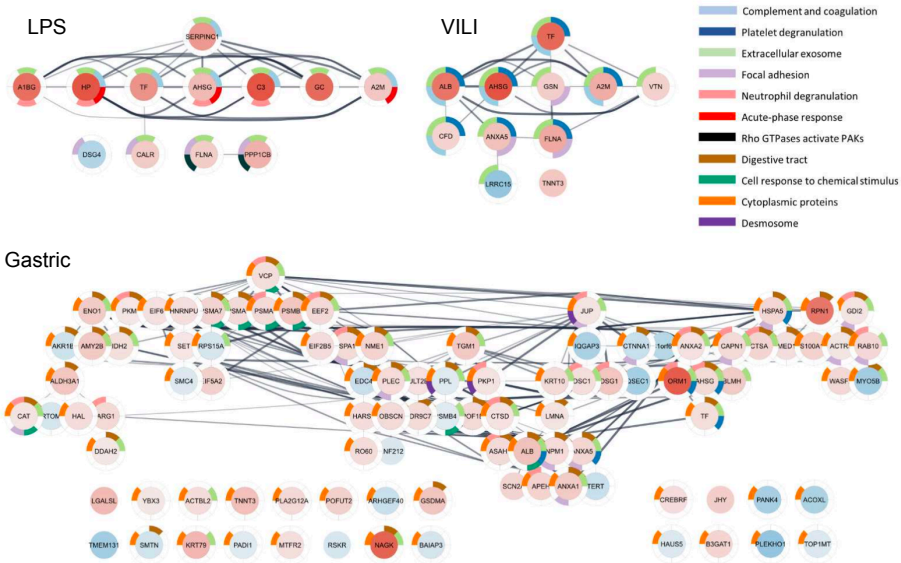


Figure 35: Protein-protein interaction network of significantly differentially expressed proteins in lung injury. Highly overexpressed proteins are shown in dark red, and underexpressed proteins in blue. Rings around proteins show how each protein maps to significantly enriched pathways within each group. The three groups are: infection (LPS), gastric aspiration (Gastric), and ventilator-induced lung injury (VILI). Adapted from Hirdman et al.[185] with permission.

Study IV

Sample Collection and Study Groups

EBP samples were collected from 230 patients at Skåne University Hospital in Skåne and divided into two groups. Patients with NSCLC awaiting surgical resection (NSCLC, n = 124) and age-matched control patients without lung cancer (Normal, n = 106). Paired tumor tissue (Tumor, n = 78) and normal adjacent tissue (NAT, n = 50) were collected intraoperatively from a subset of the NSCLC patients.

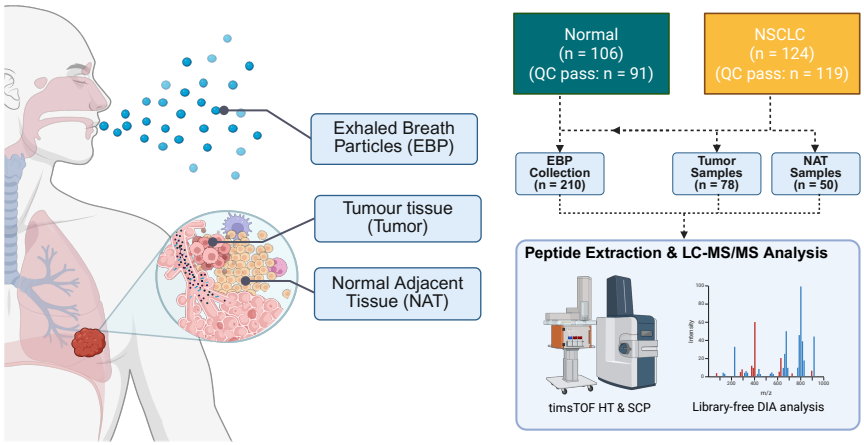


Figure 36: Experimental overview. Exhaled breath particles (EBP) were collected from 106 healthy participants (Normal) and 124 participants with non-small cell lung cancer (NSCLC). Tumor tissue and normal adjacent tissue (NAT) were furthermore collected from NSCLC participants where possible. Figure made with biorender.com.

Proteomic Remodeling in NSCLC Tumor Tissue

To establish a biological foundation for interpreting non-invasive EBP biomarker signals, we first set out to characterize the proteomic landscape of the tumor microenvironment using LC-MS/MS analysis of tumor and NAT samples. 5,450 proteins were confidently identified, and 3,192 of these were found to be significantly differentially expressed (DEPs; $q < 0.05$, $|\log_2 FC| > 0.21$), representing 59% of the total quantified lung tissue proteome. GO overrepresentation analysis of upregulated proteins in the tumor tissue revealed marked upregulation of ribonucleoprotein complex biogenesis, cytoplasmic translation, and RNA-related pathways. Among the significantly downregulated proteins, pathways related to cell-substrate adhesion, complement activation, and coagulation were identified. Using a GSEA analysis with KEGG gene sets, these findings were reinforced, showing a significant increase in Ribosome and Spliceosome activity, as well as downregulation of ECM-receptor interaction, focal adhesion, and complement and coagulation cascade proteins.

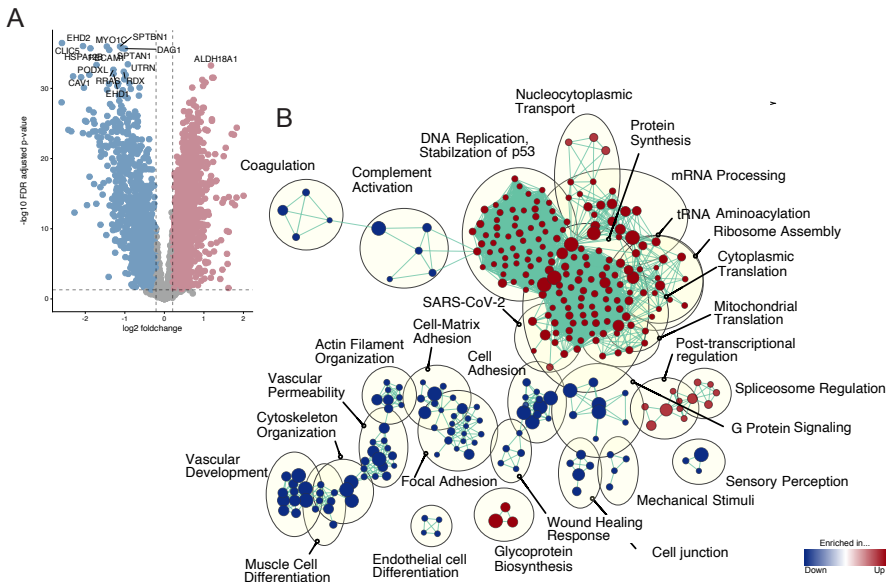


Figure 37: Proteomic differences in non-small cell lung cancer (NSCLC) tissue. A) Volcano plot of proteomic changes in NSCLC compared with normal adjacent tissue (NAT). B) Gene set enrichment analysis map of groups of enriched and downregulated pathways in NSCLC tissue. Nodes represent pathways and edges shared proteins.

EBP Proteome Composition and Cancer-Associated Signatures

We next investigated the proteomic composition of the EBP samples with LC-MS/MS. Following quality control and outlier removal, 210 EBP samples (NSCLC = 119, Normal = 91) were retained, and following peptide-level filtering, 316 proteins remained for downstream analysis.

Differential expression analysis of NSCLC versus Normal EBP samples revealed 63 DEPS ($q < 0.05$, $|\log_2 FC| > 0.17$), indicating detectable changes in the EBP proteomes of lung cancer patients. GO overrepresentation analysis of the significantly upregulated proteins identified enrichment of proteins related to cell killing, leukocyte-mediated cytotoxicity, leukocyte migration, and the humoral immune response. The downregulated proteins were strongly associated with complement activation and the humoral immune response, and GSEA KEGG analysis confirmed significant downregulation of the complement and coagulation cascade pathway, mirroring the tissue-level findings.

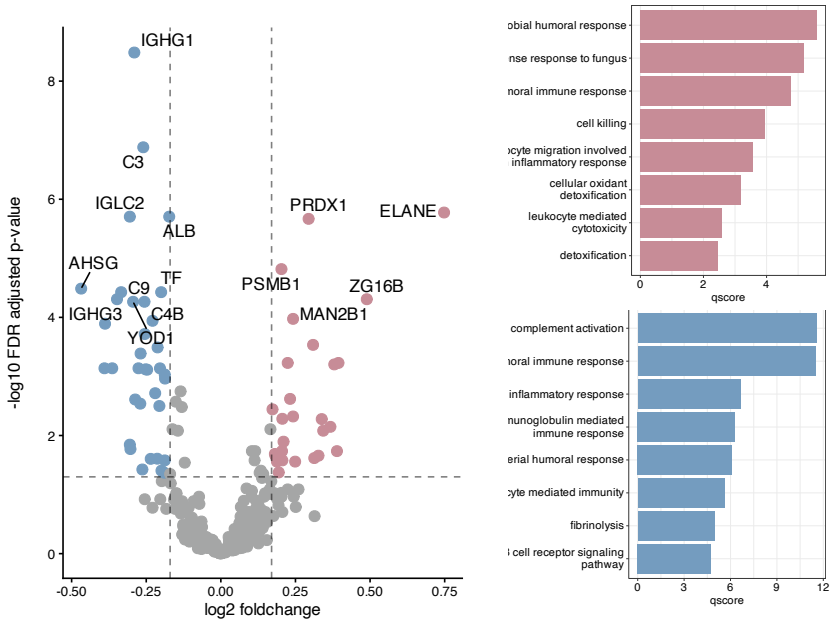


Figure 38: Protein expression in exhaled breath particles (EBP) in patients with non-small cell lung cancer (NSCLC). Volcano plot (left) of differentially expressed proteins in EBP and gene ontology (GO) overrepresentation analysis of biological processes to which up- and downregulated proteins belong. (right).

STRING protein-protein interaction network analysis of the 63 DEPs revealed a highly interconnected network with 6.2-fold more interactions than expected by chance ($p < 0.00001$). Markov Cluster Algorithm (MCL) clustering analysis resolved six distinct functional clusters of proteins. The largest cluster (n = 29) was comprised of proteins associated with blood microparticles and systemic inflammation, including neutrophil proteases (MPO, PRTN3, ELANE), complement cascade proteins (C3, CFH, CFB), acute-phase reactants (ALB, AHSG, ORM1), and protease-antiprotease balance (SERPINA1, A2M). ELANE was the most strongly differentially expressed protein in the EBP dataset and, together with MPO, LYZ, and PRTN3, formed a complex associated with neutrophil extracellular trap (NET) formation. Additional modules included a proteasome 20S subunit–centered cluster, an epithelial barrier-disruption module, an antioxidant stress-response cluster, and a lung-specific surfactant protein cluster indicative of alveolar epithelial injury as well as neutrophil aggregation proteins.

Extracellular Vesicles Enrichment of the EBP Proteome

GO cellular compartment analysis of the EBP proteome, in comparison with the tissue proteome, identified highly significant overrepresentation of secretory

granule lumen (GO:0034774; fold enrichment = 15.1), blood microparticle (GO:0072562; fold = 21.3), vacuolar lumen (GO:0005775; fold = 13.7), and primary lysosome (GO:0005766; fold = 12.5). By contrast, the tumor tissue proteome was mostly enriched for cytoplasmic and ribonucleoprotein complex terms.

Cross-referencing the identified EBP proteome with the Vesiclepedia database, which includes all known human proteins found in extracellular vesicles, revealed that 91.8% of EBP proteins were present in Vesiclepedia and as much as 98.4% of DEPs.

Integration of Tissue and EBP Proteomes

To compare the relationship between EBP and tissue proteomic overlap, we next performed an integrative analysis of the 47 patients in the study for whom matched EBP, tumor tissue, and NAT proteomes were available. 235 proteins were found across all datasets; of these, 38 were significantly differentially expressed in both EBP (NSCLC vs Normal) and tissue (Tumor vs NAT).

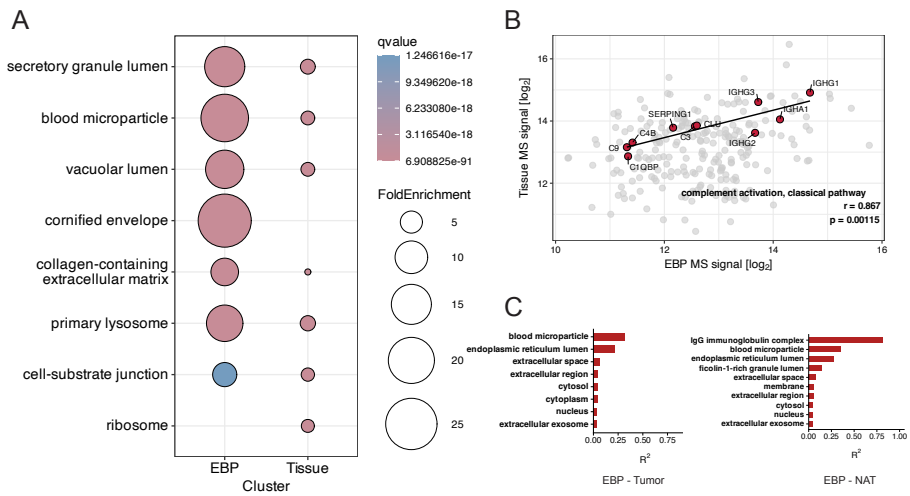


Figure 39: Correlation and differences between exhaled breath particles (EBP) and tissue in non-small cell lung cancer (NSCLC). A) Gene ontology (GO) overrepresentation analysis of cellular components of all identified proteins in EBP and tissue. B) Correlation of EBP signal intensity (x-axis) with tissue signal intensity (y-axis) for proteins belonging to the complement activation pathway. C) Top pathway correlations between EBP and tumor proteins and EBP and normal adjacent tissue (NAT).

Mean $\log_2$ protein intensity correlation for the 235 shared proteins showed a weak but significant positive association between EBP and tumor tissue (Pearson r = 0.257, $p < 0.0001$). This was somewhat higher for the 38 DEPs shared between compartments (Pearson r = 0.334, p = 0.0405). Pathway-specific analysis, however,

revealed proteins with substantially stronger correlations. Proteins annotated as part of complement activation were found to have a high correlation between EBP and tissue (Pearson $r = 0.867$, $p = 0.00115$; $n = 10$ proteins). Furthermore, analysis at the GO cellular compartments level revealed that "blood microparticle" exhibited the strongest protein expression correlation with EBP ($r = 0.593$, $R^2 = 0.351$, $p = 1.76 \times 10^{-5}$, $n = 45$ proteins), followed by endoplasmic reticulum lumen, extracellular space/region, and extracellular exosome. Performing the same analysis against the NAT proteome instead yielded nearly identical compartment rankings, with the addition of a strong correlation for IgG immunoglobulin complexes. This would suggest that EBP signals originate predominantly from secreted and vesicle-associated proteins rather than from intracellular leakage, and are accompanied by changes in the immune response in the surrounding tissue.

Development of an EBP-Based Machine Learning Classifier for NSCLC Detection

To assess the diagnostic potential of EBP proteomics for NSCLC detection, we developed a machine-learning classifier. First, the dataset was split into a discovery cohort ($n = 197$; Cancer = 111, Normal 86) and a hold-out validation set ($n = 13$, Cancer = 8, Normal = 5) made out entirely of an independent extraction batch to avoid data leakage.

First, a nested cross-validation framework (5-fold, with a 3-repeat outer loop and an inner hyperparameter-tuning loop) was set up to benchmark four SSP classifiers. These were multiclassPairs k-Top Scoring Pairs (MCP KTSP), multiclassPairs Random Forest (MCP RF), Rank-Tree Ensemble Random Forest (RTE RF), and Rank-Tree Ensemble Boost (RTE Boost). Although the average out-of-fold ROC-AUC was similar across the models, MCP RF and MCP KTSP classifiers achieved notably lower Brier scores. Calibration curve analysis further indicated near-perfect calibration for the MCP RF model, while the RTE models consistently showed poorer calibration. Based on these results, the MCP RF classifier was selected for final model development.

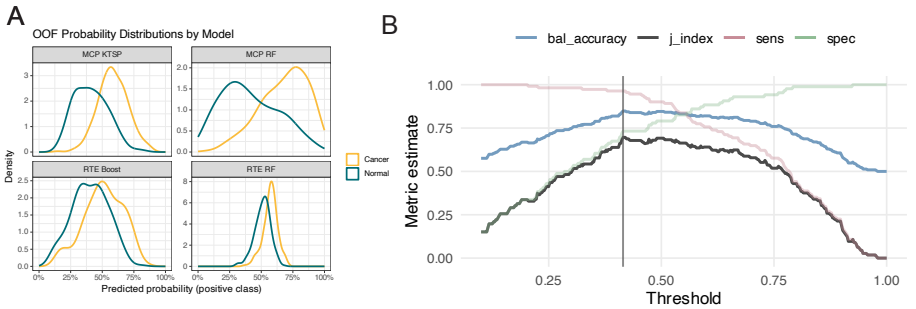


Figure 40: Benchmarking of machine learning classifiers and threshold optimization in training data. A) Out-of-fold probability distributions per machine learning classifier, with predicted probability of being a cancer sample on the x-axis and density on the y-axis. Models with a smaller area under the curves in common perform better. B) Balanced accuracy (blue), Youden's J optimized index (horizontal line), sensitivity (red line), and specificity (green line) in the training data. Youden's J identifies the threshold that maximizes the balance between sensitivity and specificity.

Next, based on the strong correlation between extracellular-related proteins and tumor tissue, a biologically informed feature selection step was added. Initial protein weights were reweighted using data from the Vesiclepedia database, giving extra emphasis to proteins identified in EVs from lung cancer cells. This slightly improved the model's ROC AUC values.

The final MCP RF model, trained on the entire discovery cohort with biologically informed reweighting, achieved an ROC AUC of 0.919 and an average precision (AP) of 0.927 on the training data.

Validation of the EBP Classifier

The finalized and locked classifier was next tested on the batch-independent hold-out validation set (n = 13), which had not been used during any phase of model training or selection, and was normalized separately to the discovery cohort.

At the optimized Youden's J threshold, the classifier achieved a final AUC of 0.750 with an AP of 0.839. All cancer patients were correctly identified, yielding a negative predictive value (NPV) of 1.00. Three of the 5 healthy controls were correctly classified, resulting in a specificity of 0.6 and a positive predictive value (PPV) of 0.8.

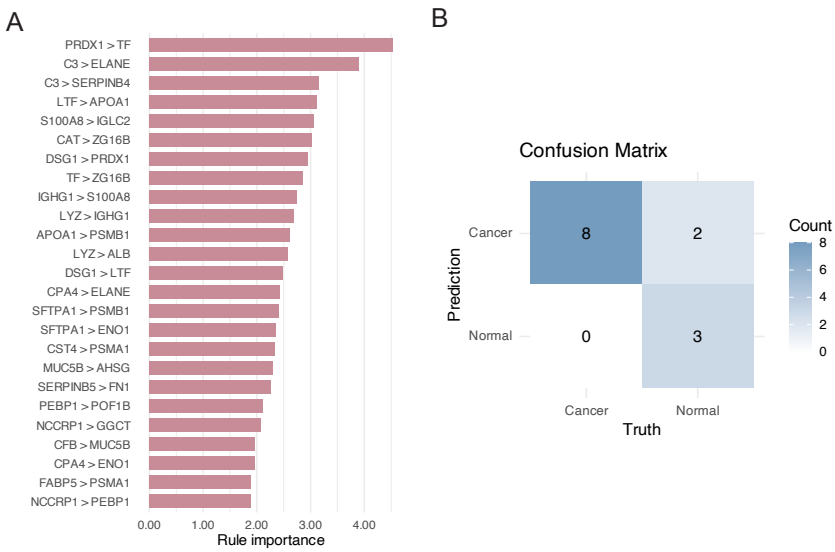


Figure 41: Results of the final locked model. A) Classifier rules ordered by importance by the random forest model B) Final confusion matrix on the independent validation set.

General Discussion

Current lung diagnostic techniques remain limited in sensitivity and specificity, or are invasive, time-consuming, and harmful due to radiation exposure and the risks associated with biopsies and BALF. Current trends in lung diagnostics are therefore moving towards non-invasive, point-of-care testing approaches[186]. The work presented in this thesis builds on a novel approach to this end by exploring proteomic signatures of exhaled breath particles. Throughout Studies I-IV, the depth of the EBP proteome expanded progressively, from approximately 110 proteins per sample in Study I to 1,923 unique proteins across the NSCLC cohort in Study IV, enabling a deeper understanding of origin and underlying biological processes compared with other methods, requiring pooling of samples to achieve comparable protein identification depth[187].

Particle flow rate as a biomarker

Consistent with previous research on EBP, increased particle concentrations were found in subjects with respiratory symptoms in Studies I-III[83,85,188]. Studies on droplet formation in film-bursting models provide a possible explanation for this: increases in surface tension lead to higher droplet concentrations[189]. Surfactants in the lung, particularly SFTPB and SFTPC, are surface-active agents that reduce the surface tension of water and were consistently and significantly underexpressed in patients with respiratory symptoms, potentially explaining the increases and changes in particle size distribution profiles identified in the study. This was further corroborated in Studies II and III, with significant increases in PFR among lung transplant recipients with PGD, as well as across all three models of lung injury in Study III.

In Studies II and III, we set out to investigate this relationship mechanistically, through combined proteomic analyses of BALF and tissue, with a focus on alveolar-capillary barrier breakdown and ECM changes. In Study II, adherens junction proteins, such as VE-cadherin, vinculin, junctional plakoglobin, and catenin-α-1, along with tight junction proteins such as zona occludens, were underexpressed in lung tissue from PGD recipients, with concordant upregulation in BALF from recipients with PGD. Tight- and adherens junctions have been found to be important for maintaining permeability, and knockout studies of junctional proteins have implicated downregulation of such proteins as a major mechanism for fluid buildup

in the lungs in ARDS[190–192]. Furthermore, an increase in total BALF protein content and in high-molecular-weight proteins was observed, providing further evidence of alveolar-capillary breakdown in lung injury. This was accompanied by the loss of AQP-5, which is normally expressed in the apical membrane of alveolar type I epithelial cells, where it serves as a principal route for water transport across the alveolar barriers and supports airway surface homeostasis[193]. It has consistently been found to be downregulated in multiple lung injury models, thereby impairing alveolar fluid clearance, potentially diluting endogenous pulmonary surfactants, and increasing surface tension[193].

Disease-specific proteomic alterations captured by EBP

While increased PFR may serve as a general marker of lung injury, proteomic alterations of EBP could serve as biomarkers of specific disease states. In Study I, a signature of acute infectious pulmonary disease was identified. Proteins involved in immune activation, acute-phase response, cell adhesion, and blood coagulation, and the main components of the RTLF, were identified as differentially expressed and could be used to distinguish and cluster the EBP proteome by disease status. COVID-19 patients exhibited significant upregulation of acute-phase proteins in EBP, including alpha-1-acid glycoprotein 1 (ORM1), alpha-1-antitrypsin (SERPINA1), and haptoglobin (HP). ORM1, sometimes referred to as orosomucoid 1, is mainly produced by hepatocytes in response to stress, but has also been reported to be produced by AT2 cells upon LPS stimulation. It was identified in almost all EBP samples and was significantly elevated in COVID-19-positive patients compared to both symptomatic but COVID-19-negative patients and healthy controls[194]. Apolipoprotein A-I (APOA1), a stress response protein with established anti-inflammatory and antiviral properties in the lung epithelium, was also significantly elevated in COVID-19-positive patients in EBP[195–198]. Interestingly, plasma proteomic studies of COVID-19 have reported decreased levels of APOA1, whereas increases in BALF have been reported to correlate with the severity of lung injury[199–202]. This discordance between the systemic and pulmonary compartments highlights the added value of sampling directly from the RTLF, where localized upregulation may precede or differ from changes detectable in peripheral blood.

In Study II, we examined proteomic alterations in PGD following lung transplantation and found distinct proteomic profiles in the EBP from affected recipients. Seven proteins were found to be significantly overexpressed in the EBP of porcine recipients with PGD, and 18 in the human EBP PGD proteome, with concordant increases in the heavy-molecular-weight proteins albumin and IgM, as well as the immune-response proteins fibrin (FGA) and ITIH4. Likely due to alveolar-capillary breakdown as discussed previously. EBP could therefore potentially be used for early detection of plasma leakage into the alveoli and prompt intervention. In Study III, three different donor lung injury models were

investigated, revealing consistent alterations in ECM proteins across all models. This was most striking in the LPS group, where ECM proteins accounted for 42.4% of all DEPs, despite only comprising 4.5% of the total identified proteome. The ECM plays a pivotal role in maintaining barrier functions and regulating the trafficking of immune cells, fluids, and molecules, and changes in ECM composition have been reported to precede inflammatory reactions[203,204]. Across all three injury models, collagen type VI alpha chains (COL6A1, COL6A2, COL6A6) and calgranulin proteins (S100A8, S100A12) were consistently altered. Collagen VI is highly expressed in the lung basement membrane and plays important roles in tissue maintenance and cell adhesion. Its proteolytic fragments have been proposed as disease markers for fibrosis, but our results suggest that COL6 degradation may be detectable even earlier, during the acute phase of lung injury[205]. The tissue proteomic data also exhibited distinct pathway activation: T-cell proliferation and complement suppression were unique to gastric aspiration, monocyte chemotaxis was characteristic of VILI, and stress-response protein activation defined the LPS-mediated injury. This was reflected in the EBP collected during mechanical ventilation. In the LPS group, acute-phase response proteins were significantly overexpressed, and in the gastric group, many digestive tract-related proteins and cytosolic proteins were identified. These findings could help lung transplant teams guide therapeutic interventions for gastric aspiration-induced injury using EBP-based proteomics detection.

Study IV revealed the highest number of differentially expressed proteins across all experiments in this thesis, reflecting improvements in the analytical pipeline and substantial differences in proteomic expression between tissue and NAT. The upregulated proteins in EBP were dominated by immune- and cytotoxicity-related signatures, including leukocyte-mediated cytotoxicity, cell killing, and the humoral immune response. Neutrophil elastase (ELANE) emerged as the most strongly differentially expressed protein and, together with myeloperoxidase (MPO), lysozyme C (LYZ), and proteinase 3 (PRTN3), formed a network associated with neutrophil extracellular trap (NET) formation[206,207]. NETs have been increasingly recognized as important contributors to tumor-promoting inflammation, and NET-associated signatures have been described in the tumor microenvironment of various cancers[152,208]. The complement and coagulation cascades were significantly downregulated in NSCLC EBP, mirroring findings from tumor tissue proteomics, where these pathways are well established as dysregulated[209,210]. This co-dysregulation was confirmed by per-patient correlation analysis of tissue and EBP, with complement activation showing a particularly strong cross-compartment correlation.

Establishment of EBP proteomics as a viable sampling method

Study I confirmed EBP sampling as a safe and effective method for RTLF sampling from the distal lungs, with no adverse events reported, even when sampling patients reporting dyspnea. Studies II and III further demonstrated the safety and efficacy of sampling EBP during mechanical ventilation[211]. The preferential capture of distal RTLF-related proteins observed in Study I and in previously published research was further extended in Study II through direct comparison of BALF proteomes with porcine and human EBP. GO enrichment analysis revealed enrichment for acute-phase and inflammatory response terms, as well as cytoskeletal and filament organization pathways, consistent with the expected composition of RTLF during lung injury[212,213].

Study IV substantially advanced the understanding of what EBP biologically represents. Cross-referencing the entire EBP proteome with Vesiclepedia, a curated database of extracellular vesicle (EV) contents, demonstrated that 91.8% of all detected EBP proteins and 98.4% of differentially expressed proteins corresponded to previously reported vesicular proteins. GO cellular compartment analysis confirmed strong enrichment for secretory granule lumen, blood microparticle, and vacuolar lumen terms. This is in line with previous findings in EBP and EBC, and cross-correlations with the tissue proteome suggest that EBP should be considered a distinct biological compartment enriched for EVs and secretory particles released into the RTLF by immune, tumor, and endothelial cells[214–216]. EVs have attracted substantial interest as disease biomarkers, but obtaining lung-tissue-derived EVs has traditionally required BALF or difficult-to-enrich liquid biopsies. EBP collection thus provides a non-invasive route for accessing this vesicular compartment directly from the lung milieu[217–220].

EBP proteomics as a diagnostic tool

The diagnostic potential of EBP proteomics was explored in Studies I and IV. In Study I, a random forest machine-learning classifier was trained on an integrated panel of 11 proteins and particle-concentration counts from 22 patients, achieving a test accuracy of 92% for distinguishing COVID-19-positive patients. In Study IV, this diagnostic approach was scaled up substantially, with the training set comprising 197 subjects and model selection performed via a rigorous nested cross-validation framework to benchmark four SSP classifiers. SSP classifiers were chosen for their ability to classify individual specimens without requiring cohort-level normalization, thereby improving robustness and facilitating clinical deployment[221–223]. The selected multiclassPairs random forest model achieved impressive sensitivity and an overall balanced accuracy of 80%.

The performance characteristics of the NSCLC classifier suggested a potential role for EBP-based screening as a pre-LDCT triage tool. In the National Lung Screening

Trial (NLST), approximately one-quarter of LDCT screens were positive, and 96.4% of these positive scans were false positives[155]. A high-sensitivity, moderate specificity triage test could thus reduce unnecessary LDCT scans, thereby limiting downstream harms, including avoidable radiation exposure, patient anxiety, and invasive follow-up procedures, and have been requested previously to this end[130].

Limitations

Several limitations should be acknowledged across the four studies in this thesis. All patient cohorts were recruited from a single center (Skåne University Hospital), which may limit the generalizability of the findings. Sample sizes, while adequate for exploratory proteomics, remain modest for clinical implementation and need to be validated in large multi-center trials. For the machine learning classifiers, test sets were very small in Studies I and IV, and the reported sensitivities and specificities should thus be interpreted with caution. The porcine models used in Studies II and III, while the closest available approximation to human data, may be influenced by species differences that do not necessarily apply to humans.

Conclusion and Future Perspectives

The work presented in this thesis expands on the use of exhaled breath particle (EBP) proteomic analysis in a variety of acute and malignant lung diseases. Sampling of exhaled breath particles represents a versatile, non-invasive platform for pulmonary diagnostics. Study I demonstrated the feasibility of using mass spectrometry-based proteomic analysis and particle production analysis in COVID-19. Potential biomarkers in this study included increases in extracellular acute-phase proteins and decreases in endogenous surfactant proteins in EBP. This study further demonstrated the potential of an integrated EBP biomarker panel comprising protein expression and particle concentrations for the diagnosis of COVID-19. Study II demonstrated the use of EBP for early warning of PGD in lung transplant recipients by measuring particle concentrations and linking increases in PFR to alveolar-capillary barrier disruption in PGD, highlighting the potential of EBP proteomics to detect this change. Study III further extended EBP and tissue proteomic analysis by investigating changes in three distinct models of acute lung injury and identifying shared ECM alterations in early acute lung injury. Study IV applied paired EBP-tumor tissue proteomics in the largest EBP lung cancer cohort to date, demonstrating EV enrichment, cross-compartment proteomic correlations, and a machine-learning classifier with diagnostic potential for triaging of screening patients.

Future studies should focus on multicenter prospective validation trials and refinement of the NSCLC EBP classifier in a screening population. Improvements and further work are needed to better understand sample variability and increase the proteomic yield from EBP membranes. Third, the development of targeted, point-of-care analytical assays for specific proteins or biomarker panels could facilitate clinical implementation and, if proven effective, enable deployment of EBP analysis in smaller health-care centers. Furthermore, the integration of EBP proteomic analysis into lung transplantation workflows represents exciting new possibilities for improving organ selection, determining which organs require which interventions, and non-invasively monitoring patients for PGD post-transplantation. Future studies should focus on detecting COL6A degradation fragments in donors as a potential marker of donor lung injury, and on multi-center trials of EBP PFR surveillance in mechanically ventilated patients post PGD.

Declaration about the use of generative AI

The following large language model-assisted tools were used in order to improve language, grammar, and spelling, as well as readability and flow: Anthropic's Claude Opus 4.6 model and Grammarly. All AI-generated suggestions were critically revised by the author. The author takes full responsibility for all written content of this thesis.

Acknowledgements

I would like to express my sincere gratitude to the many people who have supported me throughout the completion of this thesis, without whom this research project would not have been possible.

I would first like to extend my heartfelt thanks to all the patients who generously contributed their time and effort to the studies included in this work. Their selfless participation made this research possible.

I am especially grateful to my main supervisor, **Professor Sandra Lindstedt**, for your invaluable guidance, encouragement, and unwavering energy and enthusiasm throughout my doctoral studies. From the very beginning, you believed in me and generously provided the support, resources, and opportunities that made this research possible. Your knowledge, insight, and deep commitment to scientific pursuit have been a constant source of inspiration. I am profoundly thankful for your mentorship.

I would also like to thank my co-supervisors, **Franziska Olm** and **Sven Kjellström**. Their valuable input, patience, and generosity in sharing their knowledge throughout this journey. Our many meetings, the lessons I have learned from you, and your unwavering support have meant a great deal to me.

I am also deeply grateful to my co-authors and lab members, **Embla Bodén, Margareta Mittendorfer, Nicholas Bèchet, Haider Ghaidan, Anna Niroomand, Martin Stenlo, Dag Edström, Leif Pierre, Snejana Hyllén,** and **Qi Wang,** for their invaluable contributions, ideas, and support. You truly make the long days and nights in the lab both incredibly enjoyable and rewarding.

This research would not have been possible without the support of the wonderful people at BioMS. I would like to extend my deepest appreciation to **Filip Årman, Marc Isaksson** and **Melinda Rezeli** for their expertise, generosity, and invaluable assistance throughout this work.

My sincere thanks also go to my colleagues and friends during these years for their constant encouragement, perspectives, shared laughter, and companionship during this process. Your presence has made this journey lighter and far more meaningful, and I will always value the support, friendship, and memorable moments.

I would also like to thank my family for their steadfast support, interest, and belief in me throughout this journey. Your love, patience, and encouragement have been

a constant source of strength, especially during the more challenging moments. I am deeply grateful for everything you have done for me and for getting me to where I am today. Thank you.

Lastly, I would like to thank **Else Ånggård**, whom I was fortunate enough to meet during this journey. Your love, support, and presence have meant more to me than words can express. Thank you for bringing joy, strength, and comfort into my life during these years. This thesis is dedicated to you.

References

1. Van Scott, M. R., Chandler, J., Olmstead, S., Brown, J. M. & Mannie, M. Airway Anatomy, Physiology, and Inflammation. Toxic. Induction Irrit. Asthma Rhinitis Relat. Cond. 19–61 (2013) doi:10.1007/978-1-4614-9044-9_2.
2. Weibel, E. R. & Gomez, D. M. Architecture of the human lung. Use of quantitative methods establishes fundamental relations between size and number of lung structures. Science 137, 577–585 (1962).
3. Ochs, M. et al. The number of alveoli in the human lung. Am. J. Respir. Crit. Care Med. 169, 120–124 (2004).
4. Marieb, E. N. Essentials of Human Anatomy & Physiology. Essentials of human anatomy & physiology (Pearson/Benjamin Cummings, San Francisco, CA, 2009).
5. Bustamante-Marin, X. M. & Ostrowski, L. E. Cilia and Mucociliary Clearance. Cold Spring Harb. Perspect. Biol. 9, a028241 (2017).
6. Khan, Y. S. & Carey, F. J. Histology, Lung. in StatPearls (StatPearls Publishing, Treasure Island (FL), 2026).
7. Hollenhorst, M. I. & Krasteva-Christ, G. Chemosensory cells in the respiratory tract as crucial regulators of innate immune responses. J. Physiol. 601, 1555–1572 (2023).
8. Gehr, P., Bachofen, M. & Weibel, E. R. The normal human lung: ultrastructure and morphometric estimation of diffusion capacity. Respir. Physiol. 32, 121–140 (1978).
9. Maina, J. N. & West, J. B. Thin and strong! The bioengineering dilemma in the structural and functional design of the blood-gas barrier. Physiol. Rev. 85, 811–844 (2005).
10. Thuringer, M., Zent, R., Lennon, R. & Plosa, E. J. Basement Membranes in Lung Development, Disease, and Repair. Matrix Biol. J. Int. Soc. Matrix Biol. 140, 123–132 (2025).
11. Faffe, D. S. & Zin, W. A. Lung Parenchymal Mechanics in Health and Disease. Physiol. Rev. 89, 759 (2009).
12. Dobbs, L. G. Pulmonary surfactant. Annu. Rev. Med. 40, 431–446 (1989).
13. Leiby, K. L., Raredon, M. S. B. & Niklason, L. E. Bioengineering the Blood-gas Barrier. Compr. Physiol. 10, 415–452 (2020).
14. Binks, A. Pulmonary Physiology for Pre-Clinical Students. (Virginia Tech Publishing, 2022). doi:10.21061/pulmonaryphysiology.
15. Betts et al. Anatomy and Physiology 2e. (Kendall Hunt Publishing, 2022).
16. Hickey, S. M., Sankari, A. & Giwa, A. O. Mechanical Ventilation. in StatPearls (StatPearls Publishing, Treasure Island (FL), 2026).

17. Fahy, J. V. & Dickey, B. F. Airway Mucus Function and Dysfunction. N. Engl. J. Med. 363, 2233–2247 (2010).
18. Wattiez, R. & Falmagne, P. Proteomics of bronchoalveolar lavage fluid. J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci. 815, 169–178 (2005).
19. Noël-Georis, I., Bernard, A., Falmagne, P. & Wattiez, R. Database of bronchoalveolar lavage fluid proteins. J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci. 771, 221–236 (2002).
20. Weise, D. O. et al. An optimized workflow for MS-based quantitative proteomics of challenging clinical bronchoalveolar lavage fluid (BALF) samples. Clin. Proteomics 20, 14 (2023).
21. Mammen, M. J. et al. Proteomic Network Analysis of Bronchoalveolar Lavage Fluid in Ex-Smokers to Discover Implicated Protein Targets and Novel Drug Treatments for Chronic Obstructive Pulmonary Disease. Pharmaceuticals 15, 566 (2022).
22. von Bredow, C., Birrer, P. & Griese, M. Surfactant protein A and other bronchoalveolar lavage fluid proteins are altered in cystic fibrosis. Eur. Respir. J. 17, 716–722 (2001).
23. Holter, J. F., Weiland, J. E., Pacht, E. R., Gadek, J. E. & Davis, W. B. Protein permeability in the adult respiratory distress syndrome. Loss of size selectivity of the alveolar epithelium. J. Clin. Invest. 78, 1513–1522 (1986).
24. Matute-Bello, G. et al. An official American Thoracic Society workshop report: features and measurements of experimental acute lung injury in animals. Am. J. Respir. Cell Mol. Biol. 44, 725–738 (2011).
25. Chroneos, Z. C., Sever-Chroneos, Z. & Shepherd, V. L. Pulmonary surfactant: an immunological perspective. Cell. Physiol. Biochem. Int. J. Exp. Cell. Physiol. Biochem. Pharmacol. 25, 13–26 (2010).
26. Heinrich, S., Hartl, D. & Griese, M. Surfactant protein A--from genes to human lung diseases. Curr. Med. Chem. 13, 3239–3252 (2006).
27. Schürch, D., Ospina, O. L., Cruz, A. & Pérez-Gil, J. Combined and independent action of proteins SP-B and SP-C in the surface behavior and mechanical stability of pulmonary surfactant films. Biophys. J. 99, 3290–3299 (2010).
28. Leiva-Juárez, M. M., Kolls, J. K. & Evans, S. E. Lung epithelial cells: therapeutically inducible effectors of antimicrobial defense. Mucosal Immunol. 11, 21–34 (2018).
29. Corthésy, B. Multi-faceted functions of secretory IgA at mucosal surfaces. Front. Immunol. 4, 185 (2013).
30. Strunk, R. C., Eidlen, D. M. & Mason, R. J. Pulmonary alveolar type II epithelial cells synthesize and secrete proteins of the classical and alternative complement pathways. J. Clin. Invest. 81, 1419–1426 (1988).
31. Sarma, V. J., Huber-Lang, M. & Ward, P. A. Complement in lung disease. Autoimmunity 39, 387–394 (2006).
32. Vareille, M., Kieninger, E., Edwards, M. R. & Regamey, N. The Airway Epithelium: Soldier in the Fight against Respiratory Viruses. Clin. Microbiol. Rev. 24, 210–229 (2011).

33. Zareba, L., Szymanski, J., Homoncik, Z. & Czystowska-Kuzmicz, M. EVs from BALF—Mediators of Inflammation and Potential Biomarkers in Lung Diseases. Int. J. Mol. Sci. 22, 3651 (2021).
34. Holtzman, J. & Lee, H. Emerging role of extracellular vesicles in the respiratory system. Exp. Mol. Med. 52, 887–895 (2020).
35. Levänen, B. et al. Altered microRNA profiles in bronchoalveolar lavage fluid exosomes in asthmatic patients. J. Allergy Clin. Immunol. 131, 894–903 (2013).
36. Mahida, R. Y. et al. CD14-positive extracellular vesicles in bronchoalveolar lavage fluid as a new biomarker of acute respiratory distress syndrome. Am. J. Physiol. Lung Cell. Mol. Physiol. 322, L617–L624 (2022).
37. d'Alessandro, M. et al. The Effects of Interstitial Lung Diseases on Alveolar Extracellular Vesicles Profile: A Multicenter Study. Int. J. Mol. Sci. 24, 4071 (2023).
38. Carvalho, A. S. et al. Is the Proteome of Bronchoalveolar Lavage Extracellular Vesicles a Marker of Advanced Lung Cancer? Cancers 12, 3450 (2020).
39. Dweik, R. A. & Amann, A. Exhaled breath analysis: the new frontier in medical testing. J. Breath Res. 2, 030301 (2008).
40. Phillips, M. Breath Tests in Medicine. Sci. Am. 267, 74–79 (1992).
41. Lemos, F. F. B. et al. Urea breath test for Helicobacter pylori infection in adult dyspeptic patients: A meta-analysis of diagnostic test accuracy. World J. Gastroenterol. 30, 579–598 (2024).
42. Moura, P. C., Raposo, M. & Vassilenko, V. Breath volatile organic compounds (VOCs) as biomarkers for the diagnosis of pathological conditions: A review. Biomed. J. 46, 100623 (2023).
43. Altomare, D. F. et al. Exhaled volatile organic compounds identify patients with colorectal cancer. Br. J. Surg. 100, 144–150 (2013).
44. Mohitkar, A., Kokate, P., Wathore, R., Kanade, G. & Middey, A. An Overview of VOC Sampling and Detection Techniques. Res. J. Chem. Environ. 27, 132–141 (2023).
45. Gallego, E., Roca, F. J., Perales, J. F. & Guardino, X. Comparative Study of the Adsorption Performance of a Multi-Sorbent Bed (Carbotrap, Carbopack X, Carboxen 569) and a Tenax TA Adsorbent Tube for the Analysis of Volatile Organic Compounds (VOCs). Talanta 81, 916–924 (2010).
46. Kim, Y.-H. & Kim, K.-H. Extent of Sample Loss on the Sampling Device and the Resulting Experimental Biases When Collecting Volatile Fatty Acids (VFAs) in Air Using Sorbent Tubes. Anal. Chem. 85, 7818–7825 (2013).
47. Wong, G. K. S., Ng, S. J. & Webster, R. D. Quantitative Analysis of Atmospheric Volatile Organic Pollutants by Thermal Desorption Gas Chromatography Mass Spectrometry. Anal Methods 5, 219–230 (2013).
48. Oliver, K. D. et al. Sample Integrity Evaluation and EPA Method 325B Interlaboratory Comparison for Select Volatile Organic Compounds Collected Diffusively on Carbopack X Sorbent Tubes. Atmos. Environ. 163, 99–106 (2017).
49. Sun, X. & Yang, X. Volatile Organic Compounds in Normal Human Exhaled Breath: A Long Neglected Pollutant Source. WIT Trans. Built Environ. 1, 773–783 (2013).

50. Shaw, J. G. et al. Biomarkers of progression of chronic obstructive pulmonary disease (COPD). J. Thorac. Dis. 6, 1532–1547 (2014).
51. Oakley-Girvan, I. & Davis, S. W. Breath based volatile organic compounds in the detection of breast, lung, and colorectal cancers: A systematic review. Cancer Biomark. 21, 29–39 (2018).
52. Gasparri, R. et al. Diagnostic biomarkers for lung cancer prevention. J. Breath Res. 12, 027111 (2018).
53. Narendra, D., Blixt, J. & Hanania, N. A. Immunological biomarkers in severe asthma. Semin. Immunol. 46, 101332 (2019).
54. Stavropoulos, G. et al. Liver Impairment—The Potential Application of Volatile Organic Compounds in Hepatology. Metabolites 11, 618 (2021).
55. Gao, Q. & Lee, W.-Y. Urinary metabolites for urological cancer detection: a review on the application of volatile organic compounds for cancers. Am. J. Clin. Exp. Urol. 7, 232–248 (2019).
56. Scheideler, L., Manke, H. G., Schwulera, U., Inacker, O. & Hämmerle, H. Detection of nonvolatile macromolecules in breath. A possible diagnostic tool? Am. Rev. Respir. Dis. 148, 778–784 (1993).
57. Konstantinidi, E. M., Lappas, A. S., Tzortzi, A. S. & Behrakis, P. K. Exhaled Breath Condensate: Technical and Diagnostic Aspects. Sci. World J. 2015, 435160 (2015).
58. Effros, R. M. et al. A simple method for estimating respiratory solute dilution in exhaled breath condensates. Am. J. Respir. Crit. Care Med. 168, 1500–1505 (2003).
59. Papineni, R. S. & Rosenthal, F. S. The size distribution of droplets in the exhaled breath of healthy human subjects. J. Aerosol Med. Off. J. Int. Soc. Aerosols Med. 10, 105–116 (1997).
60. Jackson, A. S. et al. Comparison of Biomarkers in Exhaled Breath Condensate and Bronchoalveolar Lavage. Am. J. Respir. Crit. Care Med. 175, 222–227 (2007).
61. Malerba, M., Purghè, B., Ragnoli, B., Manfredi, M. & Baldanzi, G. Molecular profiling of exhaled breath condensate in respiratory diseases. Ann. Med. 57, 2537910 (2025).
62. Almstrand, A.-C. et al. Airway monitoring by collection and mass spectrometric analysis of exhaled particles. Anal. Chem. 81, 662–668 (2009).
63. Haslbeck, K., Schwarz, K., Hohlfeld, J. M., Seume, J. R. & Koch, W. Submicron droplet formation in the human lung. J. Aerosol Sci. 41, 429–438 (2010).
64. Almstrand, A.-C. et al. Effect of airway opening on production of exhaled particles. J. Appl. Physiol. 108, 584–588 (2010).
65. Johnson, G. R. & Morawska, L. The Mechanism of Breath Aerosol Formation. J. Aerosol Med. Pulm. Drug Deliv. 22, 229–237 (2009).
66. Holmgren, H., Ljungström, E., Almstrand, A.-C., Bake, B. & Olin, A.-C. Size distribution of exhaled particles in the range from 0.01 to 2.0 μm. J. Aerosol Sci. 41, 439–446 (2010).
67. Johnson, G. R. et al. Modality of human expired aerosol size distributions. J. Aerosol Sci. 42, 839–851 (2011).

68. Bredberg, A. et al. Exhaled Endogenous Particles Contain Lung Proteins. Clin. Chem. 58, 431–440 (2012).
69. Larsson, P. et al. Surfactant protein A and albumin in particles in exhaled air. Respir. Med. 106, 197–204 (2012).
70. Behndig, A. F., Mirgorodskaya, E., Blomberg, A. & Olin, A.-C. Surfactant Protein A in particles in exhaled air (PExA), bronchial lavage and bronchial wash - a methodological comparison. Respir. Res. 20, 214 (2019).
71. Viklund, E. et al. Current smoking alters phospholipid- and surfactant protein A levels in small airway lining fluid: An explorative study on exhaled breath. PLOS ONE 16, e0253825 (2021).
72. Kokelj, S. et al. Smoking induces sex-specific changes in the small airway proteome. Respir. Res. 22, 234 (2021).
73. Larsson, P. et al. Exhaled particles as markers of small airway inflammation in subjects with asthma. Clin. Physiol. Funct. Imaging 37, 489–497 (2017).
74. Carpaij, O. A. et al. Assessing small airways dysfunction in asthma, asthma remission and healthy controls using particles in exhaled air. ERJ Open Res. 5, 00202–02019 (2019).
75. Alahmadi, F. H., Wilkinson, M., Keevil, B., Niven, R. & Fowler, S. J. Short- and medium-term effect of inhaled corticosteroids on exhaled breath biomarkers in severe asthma. J. Breath Res. 16, (2022).
76. Soares, M. et al. Particles in exhaled air (PExA): non-invasive phenotyping of small airways disease in adult asthma. J. Breath Res. 12, 046012 (2018).
77. Kokelj, S. et al. Activation of the Complement and Coagulation Systems in the Small Airways in Asthma. Respir. Int. Rev. Thorac. Dis. 102, 621–631 (2023).
78. Lärstad, M. et al. Surfactant Protein A in Exhaled Endogenous Particles Is Decreased in Chronic Obstructive Pulmonary Disease (COPD) Patients: A Pilot Study. PloS One 10, e0144463 (2015).
79. Wadsworth, S., Sin, D. & Dorscheid, D. Clinical update on the use of biomarkers of airway inflammation in the management of asthma. J. Asthma Allergy 4, 77–86 (2011).
80. Matthay, M. A. & Zemans, R. L. The acute respiratory distress syndrome: pathogenesis and treatment. Annu. Rev. Pathol. 6, 147–163 (2011).
81. Ashbaugh, D., Bigelow, D. B., Petty, T. & Levine, B. ACUTE RESPIRATORY DISTRESS IN ADULTS. The Lancet 290, 319–323 (1967).
82. Fan, E., Brodie, D. & Slutsky, A. S. Acute Respiratory Distress Syndrome: Advances in Diagnosis and Treatment. JAMA 319, 698–710 (2018).
83. Stenlo, M. et al. Increased particle flow rate from airways precedes clinical signs of ARDS in a porcine model of LPS-induced acute lung injury. Am. J. Physiol. Lung Cell. Mol. Physiol. 318, L510–L517 (2020).
84. Holz, O., Müller, M., Carstensen, S., Olin, A.-C. & Hohlfeld, J. M. Inflammatory cytokines can be monitored in exhaled breath particles following segmental and inhalation endotoxin challenge in healthy volunteers. Sci. Rep. 12, 5620 (2022).

85. Stenlo, M. et al. Monitoring lung injury with particle flow rate in LPS- and COVID-19-induced ARDS. Physiol. Rep. 9, e14802 (2021).
86. Sherman, A. C. et al. Acute SARS-CoV-2 infection. Nat. Rev. Dis. Primer 11, 75 (2025).
87. Chan, J. F.-W. et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. The Lancet 395, 514–523 (2020).
88. Zhu, N. et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N. Engl. J. Med. 382, 727–733 (2020).
89. CDC. CDC Museum COVID-19 Timeline. Centers for Disease Control and Prevention https://www.cdc.gov/museum/timeline/covid19.html (2024).
90. Coronavirus disease (COVID-19) pandemic. https://www.who.int/europe/emergencies/situations/covid-19.
91. Guan, W. et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N. Engl. J. Med. 382, 1708–1720 (2020).
92. Radygina, L. V. & Mochalova, L. V. Factors affecting the severity of COVID-19 and the development of complications. Microbiol. Indep. Res. J. MIR J. 10, 20–38 (2023).
93. Yang, H. & Rao, Z. Structural biology of SARS-CoV-2 and implications for therapeutic development. Nat. Rev. Microbiol. 19, 685–700 (2021).
94. Hoffmann, M. et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell 181, 271-280.e8 (2020).
95. Ziegler, C. G. K. et al. SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. Cell 181, 1016-1035.e19 (2020).
96. Jacobs, J. L. et al. Plasma SARS-CoV-2 RNA Levels as a Biomarker of Lower Respiratory Tract SARS-CoV-2 Infection in Critically Ill Patients With COVID-19. J. Infect. Dis. 226, 2089–2094 (2022).
97. Hagman, K. et al. Prevalence and clinical relevance of viraemia in viral respiratory tract infections: a systematic review. Lancet Microbe 6, 100967 (2025).
98. Hadjadj, J. et al. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. Science 369, 718–724 (2020).
99. Ramlall, V. et al. Immune complement and coagulation dysfunction in adverse outcomes of SARS-CoV-2 infection. Nat. Med. 26, 1609–1615 (2020).
100. Pfeiler, S., Massberg, S. & Engelmann, B. Biological basis and pathological relevance of microvascular thrombosis. Thromb. Res. 133 Suppl 1, S35-37 (2014).
101. Fragkou, P. C. et al. ESCMID COVID-19 guidelines: diagnostic testing for SARS-CoV-2. Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis. 28, 812–822 (2022).
102. Chu, V. T. et al. Comparison of Home Antigen Testing With RT-PCR and Viral Culture During the Course of SARS-CoV-2 Infection. JAMA Intern. Med. 182, 701–709 (2022).

103. Hirabayashi, E. et al. Comparison of diagnostic accuracy of rapid antigen tests for COVID-19 compared to the viral genetic test in adults: a systematic review and meta-analysis. JBI Evid. Synth. 22, 1939–2002 (2024).
104. Tsang, N. N. Y. et al. Diagnostic performance of different sampling approaches for SARS-CoV-2 RT-PCR testing: a systematic review and meta-analysis. Lancet Infect. Dis. 21, 1233–1245 (2021).
105. Baron, A. et al. Bronchoalveolar Lavage in Patients with COVID-19 with Invasive Mechanical Ventilation for Acute Respiratory Distress Syndrome. Ann. Am. Thorac. Soc. 18, 723–726 (2021).
106. Christie, J. D., Raemdonck, D. V. & Fisher, A. J. Lung Transplantation. N. Engl. J. Med. 391, 1822–1836 (2024).
107. Arjuna, A., Olson, M. T. & Walia, R. Current trends in candidate selection, contraindications, and indications for lung transplantation. J. Thorac. Dis. 13, 6514–6527 (2021).
108. Orens, J. B. et al. International guidelines for the selection of lung transplant candidates: 2006 update--a consensus report from the Pulmonary Scientific Council of the International Society for Heart and Lung Transplantation. J. Heart Lung Transplant. Off. Publ. Int. Soc. Heart Transplant. 25, 745–755 (2006).
109. Bunnik, E. M. Ethics of allocation of donor organs. Curr. Opin. Organ Transplant. 28, 192–196 (2023).
110. Valapour, M. et al. OPTN/SRTR 2023 Annual Data Report: Lung. Am. J. Transplant. Off. J. Am. Soc. Transplant. Am. Soc. Transpl. Surg. 25, S422–S489 (2025).
111. Lesko, M. B. & Angel, L. F. Organ Donation, the Non-Perfect Lung Donor, and Variability in Conversion to Transplant. Clin. Chest Med. 44, 69–75 (2023).
112. Pierre, A. F., Sekine, Y., Hutcheon, M. A., Waddell, T. K. & Keshavjee, S. H. Marginal donor lungs: a reassessment. J. Thorac. Cardiovasc. Surg. 123, 421–427; discussion, 427–428 (2002).
113. Christie, I. G. et al. National Trends in Extended Criteria Donor Utilization and Outcomes for Lung Transplantation. Ann. Thorac. Surg. 111, 421–426 (2021).
114. Van Raemdonck, D. et al. Lung Donor Selection and Management. Proc. Am. Thorac. Soc. 6, 28–38 (2009).
115. Steen, S. et al. Transplantation of lungs from a non-heart-beating donor. Lancet 357, 825–829 (2001).
116. Steen, S. et al. First Human Transplantation of a Nonacceptable Donor Lung After Reconditioning Ex Vivo. Ann. Thorac. Surg. 83, 2191–2194 (2007).
117. Nakata, K., Alderete, I. S., Hughes, B. A. & Hartwig, M. G. Ex vivo lung perfusion: recent advancements and future directions. Front. Immunol. 16, 1513546 (2025).
118. Van Raemdonck, D., Ceulemans, L. J., Van Beersel, D. & Neyrinck, A. Current achievements and future applications of ex vivo lung perfusion; where do we go from here? J. Thorac. Cardiovasc. Surg. 165, 920–924 (2023).
119. Peltz, M. A narrative review of the impact of donor factors and selection criteria on outcomes after lung transplantation. Curr. Chall. Thorac. Surg. 5, (2023).

120. McCaig, A., Sage, A., Keshavjee, S. & Liu, M. Biomarkers for human donor lung assessment during ex vivo lung perfusion. J. Heart Lung Transplant. Off. Publ. Int. Soc. Heart Transplant. 45, 300–308 (2026).
121. McQuiston, A. et al. Set Up for Failure: Pre-Existing Autoantibodies in Lung Transplant. Front. Immunol. 12, 711102 (2021).
122. Diamond, J. M. et al. Clinical risk factors for primary graft dysfunction after lung transplantation. Am. J. Respir. Crit. Care Med. 187, 527–534 (2013).
123. Snell, G. I. et al. Report of the ISHLT Working Group on Primary Lung Graft Dysfunction, part I: Definition and grading-A 2016 Consensus Group statement of the International Society for Heart and Lung Transplantation. J. Heart Lung Transplant. Off. Publ. Int. Soc. Heart Transplant. 36, 1097–1103 (2017).
124. Porteous, M. K., Diamond, J. M. & Christie, J. D. Primary graft dysfunction: lessons learned about the first 72 h after lung transplantation. Curr. Opin. Organ Transplant. 20, 506–514 (2015).
125. Gelman, A. E. et al. Report of the ISHLT Working Group on Primary Lung Graft Dysfunction Part III: Mechanisms: A 2016 Consensus Group Statement of the International Society for Heart and Lung Transplantation. J. Heart Lung Transplant. Off. Publ. Int. Soc. Heart Transplant. 36, 1114–1120 (2017).
126. Shepherd, H. M. et al. Updated Views on Neutrophil Responses in Ischemia-Reperfusion Injury. Transplantation 106, 2314–2324 (2022).
127. Jennekens, J. et al. Primary Graft Dysfunction in Lung Transplantation: An Overview of the Molecular Mechanisms. Int. J. Mol. Sci. 26, 6776 (2025).
128. Hunt, M. L. & Cantu, E. Primary Graft Dysfunction After Lung Transplantation. Curr. Opin. Organ Transplant. 28, 180–186 (2023).
129. Bos, S., Vos, R., Van Raemdonck, D. E. & Verleden, G. M. Survival in adult lung transplantation: where are we in 2020? Curr. Opin. Organ Transplant. 25, 268–273 (2020).
130. Hendriks, L. E. L. et al. Non-small-cell lung cancer. Nat. Rev. Dis. Primer 10, 71 (2024).
131. Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA. Cancer J. Clin. 74, 229–263 (2024).
132. Alberg, A. J., Brock, M. V., Ford, J. G., Samet, J. M. & Spivack, S. D. Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 143, e1S-e29S (2013).
133. Lariscy, J. T., Hummer, R. A. & Rogers, R. G. Lung cancer mortality among never-smokers in the United States: estimating smoking-attributable mortality with nationally representative data. Ann. Epidemiol. 45, 5–11 (2020).
134. Gazdar, A. F. Should we continue to use the term non-small-cell lung cancer? Ann. Oncol. 21, vii225–vii229 (2010).

135. Saigí, M. et al. Biological and clinical perspectives of the actionable gene fusions and amplifications involving tyrosine kinase receptors in lung cancer. Cancer Treat. Rev. 109, 102430 (2022).
136. Scheffler, M. et al. K-ras Mutation Subtypes in NSCLC and Associated Co-occuring Mutations in Other Oncogenic Pathways. J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer 14, 606–616 (2019).
137. George, J. et al. Comprehensive genomic profiles of small cell lung cancer. Nature 524, 47–53 (2015).
138. Kulhavy, J. et al. Impact of Baseline β-Catenin Comutations on Prognosis in EGFR-Mutant Lung Cancer. JCO Precis. Oncol. 9, e2400771 (2025).
139. Kim, Y. et al. An oncogenic CTNNB1 mutation is predictive of post-operative recurrence-free survival in an EGFR-mutant lung adenocarcinoma. PloS One 18, e0287256 (2023).
140. Collisson, E. A. et al. Comprehensive molecular profiling of lung adenocarcinoma. Nature 511, 543–550 (2014).
141. Ding, L. et al. Somatic mutations affect key pathways in lung adenocarcinoma. Nature 455, 1069–1075 (2008).
142. Altorki, N. K. et al. The lung microenvironment: an important regulator of tumour growth and metastasis. Nat. Rev. Cancer 19, 9–31 (2019).
143. Hanahan, D. & Coussens, L. M. Accessories to the crime: functions of cells recruited to the tumor microenvironment. Cancer Cell 21, 309–322 (2012).
144. Kargl, J. et al. Neutrophils dominate the immune cell composition in non-small cell lung cancer. Nat. Commun. 8, 14381 (2017).
145. Nazareth, M. R. et al. Characterization of human lung tumor-associated fibroblasts and their ability to modulate the activation of tumor-associated T cells. J. Immunol. 178, 5552–5562 (2007).
146. Wang, L. et al. Cancer-associated fibroblasts enhance metastatic potential of lung cancer cells through IL-6/STAT3 signaling pathway. Oncotarget 8, 76116–76128 (2017).
147. Salmon, H. et al. Matrix architecture defines the preferential localization and migration of T cells into the stroma of human lung tumors. J. Clin. Invest. 122, 899–910 (2012).
148. Gao, Y. et al. LKB1 inhibits lung cancer progression through lysyl oxidase and extracellular matrix remodeling. Proc. Natl. Acad. Sci. U. S. A. 107, 18892–18897 (2010).
149. Stawowczyk, M. et al. Matrix Metalloproteinase 14 promotes lung cancer by cleavage of Heparin-Binding EGF-like Growth Factor. Neoplasia 19, 55–64 (2017).
150. Ferrara, N., Gerber, H.-P. & LeCouter, J. The biology of VEGF and its receptors. Nat. Med. 9, 669–676 (2003).
151. Kargl, J. et al. Neutrophils dominate the immune cell composition in non-small cell lung cancer. Nat. Commun. 8, 14381 (2017).
152. Chen, Y. et al. The role of neutrophil extracellular traps in cancer progression, metastasis and therapy. Exp. Hematol. Oncol. 11, 99 (2022).

153. Hu, S., Yan, C., Tian, Y. & Sun, W. Neutrophils in non-small cell lung cancer and immunotherapy with PD-1/PD-L1 inhibitors. J. Transl. Med. 23, 1313 (2025).
154. Woodard, G. A., Jones, K. D. & Jablons, D. M. Lung Cancer Staging and Prognosis. Cancer Treat. Res. 170, 47–75 (2016).
155. The National Lung Screening Trial Research Team. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. N. Engl. J. Med. 365, 395–409 (2011).
156. De Koning, H. J. et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. N. Engl. J. Med. 382, 503–513 (2020).
157. Broberg, E. et al. Mechanically ventilated patients exhibit decreased particle flow in exhaled breath as compared to normal breathing patients. ERJ Open Res. 6, 00198–02019 (2020).
158. Gundry, R. L. et al. Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-Up Proteomics Workflow. Curr. Protoc. Mol. Biol. Ed. Frederick M Ausubel Al CHAPTER, Unit10.25 (2009).
159. Ilavenil, S. et al. Removal of SDS from biological protein digests for proteomic analysis by mass spectrometry. Proteome Sci. 14, 11 (2016).
160. Zougman, A., Selby, P. J. & Banks, R. E. Suspension trapping (STrap) sample preparation method for bottom-up proteomics analysis. Proteomics 14, 1006–1000 (2014).
161. Glatter, T. et al. Large-Scale Quantitative Assessment of Different In-Solution Protein Digestion Protocols Reveals Superior Cleavage Efficiency of Tandem Lys-C/Trypsin Proteolysis over Trypsin Digestion. J. Proteome Res. 11, 5145–5156 (2012).
162. Dupree, E. J. et al. A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of This Field. Proteomes 8, 14 (2020).
163. Sinha, A. & Mann, M. A beginner's guide to mass spectrometry–based proteomics. The Biochemist 42, 64–69 (2020).
164. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat. Methods 17, 41–44 (2020).
165. Krasny, L. & H. Huang, P. Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology. Mol. Omics 17, 29–42 (2021).
166. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat. Protoc. 11, 2301–2319 (2016).
167. Frankenfield, A. M., Ni, J., Ahmed, M. & Hao, L. Protein Contaminants Matter: Building Universal Protein Contaminant Libraries for DDA and DIA Proteomics. J. Proteome Res. 21, 2104–2113 (2022).
168. Koopmans, F., Li, K. W., Klaassen, R. V. & Smit, A. B. MS-DAP Platform for Downstream Data Analysis of Label-Free Proteomics Uncovers Optimal Workflows in Benchmark Data Sets and Increased Sensitivity in Analysis of Alzheimer's Biomarker Data. J. Proteome Res. 22, 374–386 (2023).

169. Goeminne, L. J. E., Sticker, A., Martens, L., Gevaert, K. & Clement, L. MSqRob Takes the Missing Hurdle: Uniting Intensity- and Count-Based Proteomics. Anal. Chem. 92, 6278–6287 (2020).
170. Zhang, X. et al. Proteome-wide identification of ubiquitin interactions using UbIA-MS. Nat. Protoc. 13, 530–550 (2018).
171. Degnan, D. J. et al. pmartR 2.0: A Quality Control, Visualization, and Statistics Pipeline for Multiple Omics Datatypes. J. Proteome Res. 22, 570–576 (2023).
172. Sticker, A., Goeminne, L., Martens, L. & Clement, L. Robust Summarization and Inference in Proteome-wide Label-free Quantification. Mol. Cell. Proteomics MCP 19, 1209–1219 (2020).
173. Karpievitch, Y. V., Nikolic, S. B., Wilson, R., Sharman, J. E. & Edwards, L. M. Metabolomics data normalization with EigenMS. PloS One 9, e116221 (2014).
174. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 20, 296 (2019).
175. Doncheva, N. T. et al. Cytoscape stringApp 2.0: Analysis and Visualization of Heterogeneous Biological Networks. J. Proteome Res. 22, 637–646 (2023).
176. Zhu, L. et al. CellWhere: graphical display of interaction networks organized on subcellular localizations. Nucleic Acids Res. 43, W571–W575 (2015).
177. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS J. Integr. Biol. 16, 284–287 (2012).
178. Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the PANTHER classification system. Nat. Protoc. 8, 1551–1566 (2013).
179. Petrov, P. B., Considine, J. M., Izzi, V. & Naba, A. Matrisome AnalyzeR – a suite of tools to annotate and quantify ECM molecules in big datasets across organisms. J. Cell Sci. 136, jcs261255 (2023).
180. Chitti, S. V. et al. Vesiclepedia 2024: an extracellular vesicles and extracellular particles repository. Nucleic Acids Res. 52, D1694–D1698 (2024).
181. Kuhn, M. Building Predictive Models in R Using the caret Package. J. Stat. Softw. 28, 1–26 (2008).
182. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682 (2012).
183. Hirdman, G. et al. Proteomic characteristics and diagnostic potential of exhaled breath particles in patients with COVID-19. Clin. Proteomics 20, 13 (2023).
184. Niroomand, A. et al. Proteomic Analysis of Primary Graft Dysfunction in Porcine Lung Transplantation Reveals Alveolar-Capillary Barrier Changes Underlying the High Particle Flow Rate in Exhaled Breath. Transpl. Int. Off. J. Eur. Soc. Organ Transplant. 37, 12298 (2024).
185. Hirdman, G. et al. Unraveling Molecular and Functional Responses Across 3 Lung Injury Models to Expand the Donor Lung Pool. Transplantation 109, 1166–1174 (2025).

186. Al-Hasan, T. M., Noorizadeh, M., Bensaali, F., Meskin, N. & Ait Hssain, A. Current trends and future orientation in diagnosing lung pathologies: A systematic survey. Intell. Med. 5, 23–36 (2025).
187. Yoo, E. J., Kim, J. S., Stransky, S., Spivack, S. & Sidoli, S. Advances in proteomics methods for the analysis of exhaled breath condensate. Mass Spectrom. Rev. 43, 713–722 (2024).
188. Edwards, D. A. et al. Exhaled aerosol increases with COVID-19 infection, age, and obesity. Proc. Natl. Acad. Sci. U. S. A. 118, e2021830118 (2021).
189. Holmgren, H. & Ljungström, E. Influence of film dimensions on film droplet formation. J. Aerosol Med. Pulm. Drug Deliv. 25, 47–53 (2012).
190. Herwig, M. C., Tsokos, M., Hermanns, M. I., Kirkpatrick, C. J. & Müller, A. M. Vascular endothelial cadherin expression in lung specimens of patients with sepsis-induced acute respiratory distress syndrome and endothelial cell cultures. Pathobiol. J. Immunopathol. Mol. Cell. Biol. 80, 245–251 (2013).
191. LaFemina, M. J. et al. Claudin-18 Deficiency Results in Alveolar Barrier Dysfunction and Impaired Alveologenesis in Mice. Am. J. Respir. Cell Mol. Biol. 51, 550–558 (2014).
192. Li, G. et al. Knockout mice reveal key roles for claudin 18 in alveolar barrier properties and fluid homeostasis. Am. J. Respir. Cell Mol. Biol. 51, 210–222 (2014).
193. Yadav, E., Yadav, N., Hus, A. & Yadav, J. S. Aquaporins in lung health and disease: Emerging roles, regulation, and clinical implications. Respir. Med. 174, 106193 (2020).
194. Crestani, B. et al. Inducible expression of the alpha1-acid glycoprotein by rat and human type II alveolar epithelial cells. J. Immunol. 160, 4596–4605 (1998).
195. Catapano, A. L., Pirillo, A., Bonacina, F. & Norata, G. D. HDL in innate and adaptive immunity. Cardiovasc. Res. 103, 372–383 (2014).
196. Lee, E. hee et al. Overexpression of apolipoprotein A1 in the lung abrogates fibrosis in experimental silicosis. PloS One 8, e55827 (2013).
197. Georgila, K., Vyrla, D. & Drakos, E. Apolipoprotein A-I (ApoA-I), Immunity, Inflammation and Cancer. Cancers 11, 1097 (2019).
198. Gordon, S. M., Hofmann, S., Askew, D. S. & Davidson, W. S. High density lipoprotein: it's not just about lipid transport anymore. Trends Endocrinol. Metab. TEM 22, 9–15 (2011).
199. Mehrani, H., Ghanei, M., Aslani, J. & Golmanesh, L. Bronchoalveolar lavage fluid proteomic patterns of sulfur mustard-exposed patients. Proteomics Clin. Appl. 3, 1191–1200 (2009).
200. Nukui, Y. et al. Identification of apolipoprotein A-I in BALF as a biomarker of sarcoidosis. Sarcoidosis Vasc. Diffuse Lung Dis. Off. J. WASOG 35, 5–15 (2018).
201. Schmelter, F. et al. Metabolic and Lipidomic Markers Differentiate COVID-19 From Non-Hospitalized and Other Intensive Care Patients. Front. Mol. Biosci. 8, 737039 (2021).
202. Shen, B. et al. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. Cell 182, 59-72.e15 (2020).

203. Sutherland, T. E., Dyer, D. P. & Allen, J. E. The extracellular matrix and the immune system: A mutually dependent relationship. Science 379, eabp8964 (2023).
204. Shimshoni, E. et al. Distinct extracellular-matrix remodeling events precede symptoms of inflammation. Matrix Biol. J. Int. Soc. Matrix Biol. 96, 47–68 (2021).
205. Cescon, M., Gattazzo, F., Chen, P. & Bonaldo, P. Collagen VI at a glance. J. Cell Sci. 128, 3525–3531 (2015).
206. Urban, C. F. et al. Neutrophil extracellular traps contain calprotectin, a cytosolic protein complex involved in host defense against Candida albicans. PLoS Pathog. 5, e1000639 (2009).
207. Wang, H. et al. Neutrophil extracellular traps in homeostasis and disease. Signal Transduct. Target. Ther. 9, 235 (2024).
208. Yang, L.-Y. et al. Increased neutrophil extracellular traps promote metastasis potential of hepatocellular carcinoma via provoking tumorous inflammatory response. J. Hematol. Oncol.J Hematol Oncol 13, 3 (2020).
209. Zhao, P. et al. The imbalance in the complement system and its possible physiological mechanisms in patients with lung cancer. BMC Cancer 19, 201 (2019).
210. Tang, Q., Zhang, H., Kong, M., Mao, X. & Cao, X. Hub genes and key pathways of non-small lung cancer identified using bioinformatics. Oncol. Lett. 16, 2344–2354 (2018).
211. Broberg, E. et al. Mechanically ventilated patients exhibit decreased particle flow in exhaled breath as compared to normal breathing patients. ERJ Open Res. 6, (2020).
212. Bhargava, M. et al. Proteomic profiles in acute respiratory distress syndrome differentiates survivors from non-survivors. PloS One 9, e109713 (2014).
213. Nguyen, E. V. et al. Proteomic profiling of bronchoalveolar lavage fluid in critically ill patients with ventilator-associated pneumonia. PloS One 8, e58782 (2013).
214. Bano, A. et al. Extraction and characterization of exosomes from the exhaled breath condensate and sputum of lung cancer patients and vulnerable tobacco consumers-potential noninvasive diagnostic biomarker source. J. Breath Res. 18, (2024).
215. Mitchell, M. I. et al. Exhaled breath condensate contains extracellular vesicles (EVs) that carry miRNA cargos of lung tissue origin that can be selectively purified and analyzed. J. Extracell. Vesicles 13, e12440 (2024).
216. Östling, J. et al. A novel non-invasive method allowing for discovery of pathologically relevant proteins from small airways. Clin. Proteomics 19, 20 (2022).
217. Wang, N. et al. Circulating exosomes contain protein biomarkers of metastatic non-small-cell lung cancer. Cancer Sci. 109, 1701–1709 (2018).
218. Carreca, A. P. et al. Extracellular Vesicles in Lung Cancer: Implementation in Diagnosis and Therapeutic Perspectives. Cancers 16, 1967 (2024).
219. Lee, S. E. et al. Genomic profiling of extracellular vesicle-derived DNA from bronchoalveolar lavage fluid of patients with lung adenocarcinoma. Transl. Lung Cancer Res. 10, (2021).
220. Carvalho, A. S. et al. Is the Proteome of Bronchoalveolar Lavage Extracellular Vesicles a Marker of Advanced Lung Cancer? Cancers 12, 3450 (2020).

221. Staaf, J. et al. RNA sequencing-based single sample predictors of molecular subtype and risk of recurrence for clinical assessment of early-stage breast cancer. NPJ Breast Cancer 8, 94 (2022).
222. Cirenajwis, H., Lauss, M., Planck, M., Vallon-Christersson, J. & Staaf, J. Performance of gene expression-based single sample predictors for assessment of clinicopathological subgroups and molecular subtypes in cancers: a case comparison study in non-small cell lung cancer. Brief. Bioinform. 21, 729–740 (2020).
223. Eriksson, P. et al. A comparison of rule-based and centroid single-sample multiclass predictors for transcriptomic classification. Bioinformatics 38, 1022–1029 (2022).

Printed by Media-Tryck, Lund 2026 NORDIC SWAN ECOLABEL 3041 0903

About the author

Gabriel Hirdman became involved in research during his medical studies at Lund University, where he received his MD in 2021 before continuing directly into doctoral work. His PhD focuses on the proteomic analysis of exhaled breath particles and lung tissue, with a particular interest in bioinformatics. Through this work, he aims to deepen understanding of pulmonary disease and contribute to the development of better diagnostic methods."

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Clinical Sciences

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2026:85
ISBN 978-91-8021-883-2
ISSN 1652-8220