# HOST BIOMARKERS OF RESPIRATORY INFECTION

# CHARACTERIZATION OF CXCL10 AS A BIOMARKER OF RESPIRATORY TRACT INFECTIONS DETECTABLE BY OPEN-SOURCE LATERAL FLOW IMMUNOASSAY

By DAYNA MIKKELSEN, BSC. (HONS)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Master of Science

## Descriptive Note

McMaster University Master of Science 2022, Medical Sciences, Hamilton Ontario

Title: Characterization of CXCL10 as a Biomarker of Respiratory Tract Infections Detectable by Open-Source Lateral Flow Immunoassay

Author: Dayna Mikkelsen

Supervisor: Dr. Jeremy Hirota

Number of Pages: xii, 72

## Lay Abstract

Respiratory tract infections are a leading cause of death and one of the main reasons to seek primary care. Both historically and in the present day, respiratory tract infections remain a massive socioeconomic burden. Current diagnostics fail to quickly identify a respiratory tract infection's etiology, and prognosis, leading to suboptimal patient care and the over prescription of antibiotics. Advanced tools used in academia and research, including next-generation -omics sequencing and metagenomics, have capabilities to identify all nucleic acid material in a sample - including host RNA- which offers potential to improve the diagnosing of respiratory tract infections. However, these technologies have not been integrated into routine care due to economic, technical, and logistical barriers. We explored host RNA (transcriptomics), looking at antiviral interferon-stimulated genes for their potential as a biomarker of viral infection amenable to point-of-care testing platforms from non-invasive sample types.

# Abstract

**Background**: Respiratory tract infections are responsible for millions of deaths annually. Interferon-stimulated genes (ISGs) play a significant role in fighting off viral respiratory tract infections in the antiviral defence system. Measuring extracellular protein products of ISGs could be potential biomarkers of viral infection. Although, the feasibility and performance of ISGs as functional and robust clinical biomarkers from a non-invasive sample format remains unknown.

**Methods**: Three ISGs, *CXCL10*, *CXCL11*, and *TNFSF10*, were examined in *in-vivo* and *in-vitro* gene expression datasets (RNA-sequencing and microarray) infected with common respiratory tract infections (Rhinovirus, Respiratory syncytial virus, influenza A and SARS-CoV-2) samples and compared to negative controls. Using qualitative selection criteria of 1) elevated presence in at least one dataset with viral infection, 2) secreted protein product, and 3) commercially available antibodies for detection, *CXCL10*, *CXCL11* and *TNFSF10* gene expression levels were assessed. A correlation analysis was performed with SARS-CoV-2 infection severity and gene expression kinetics. CXCL10 was subsequently validated at the protein level in saliva as a prerequisite for developing a host-response LFA.

**Results**: *CXCL10* and *CXCL11* upregulation were positively correlated with RSV compared to control ($p < 0.05$). *CXCL10/CXCL11/TNFSF10* were not different between samples collected from RV infected subjects relative to controls ($p > 0.05$). No significant association was found with influenza A for all three genes. *CXCL10/CXCL11/TNFSF10* upregulation was positively correlated with SARS-CoV-2 infection compared to control ($p < 0.001$). *CXCL10* expression correlated with COVID-19 viral load. CXCL10 was chosen as a lead biomarker candidate based on these analyses that included different virus infections, time-courses, and measures of severity. CXCL10 was not detected at the protein level in healthy saliva but was elevated in saliva from COVID-19 patients. A CXCL10 LFA was developed with a sensitivity of 2 ng/ml in a buffer and artificial saliva.

**Conclusion:** We establish and validate the potential of developing rapid test techniques to examine host immune response from a bioinformatic approach to developing a prototype rapid test with capabilities to be used in point-of-care settings.

# Acknowledgments

As my graduate degree ends, I am very thankful for the experience and opportunity McMaster and the Firestone Institute for Respiratory Health provided.

First and foremost, I would like to thank my Supervisor, Dr. Jeremy Hirota. Since starting at McMaster, Dr. Jeremy Hirota has been an incredible mentor, supervisor, and friend as we navigated figuring out what a master's project could be like during a pandemic. His scientific guidance and lessons will carry with me in my future directions; this project would not have been possible without his support.

I would also like to thank my committee members, Dr. Kjetil Ask and Dr. Matthew Miller, for encouraging critical thinking and providing mentorship and guidance throughout my master's project. In addition, Dr. Andrew Doxey, who was a great co-supervisor and mentor throughout this project who brought his team and positive outlook to the presented thesis.

Lastly, I would like to thank my parents, brother, and Jake, who have helped in countless ways, supported me through pursuing graduate school and always encouraged my passion for science. Also, I can't forget to mention Oak Ridge's Public Library, which became my second home during the writing portion of this thesis.

# Table of Contents

# List of Figures



# List of Tables

# List of Abbreviations and Symbols

ACTB – Actin Beta

CARD – Comprehensive Antibiotic Resistance Database

CIHR – Canadian Institutes Health Research

Ct – Cycle threshold

CT- Computed tomography

CXCL10 – Interferon-γ inducible protein 10 kDa (IP-10)

CXCL11 – C-X-C motif chemokine ligand 11

DAMPS – Damage associated molecular patterns

DVT – Deep Vein Thrombosis

FDA – Food and Drug Administration

FluA – Influenza A

FluB – Influenza B

GADPH – Glyceraldehyde 3-phosphate dehydrogenase

HCG – Human chorionic gonadotropin

IFN-γ – Interferon gamma

IgG - Immunoglobulin G

IL-6 – Interleukin 6

ISG – Interferon Stimulated Gene

IVD – In-vitro diagnostic

L2M – Lab2Market Program

LDT – Laboratory Developed Test

LFA – Lateral flow assay

MRI - Magnetic Resonance Imaging

NCBI GEO – National Center for Biotechnology Information Gene Expression Omnibus

NGS – Next generation sequencing

NPS – Nasopharyngeal Swab

NSERC – National Science and Engineering Research Council

PAMPS – Pathogen Associated Molecular Patterns

PCR – Polymerase Chain Reaction

PRR – Pattern recognition receptor

RMA– Robust multiarray average

RNA – Ribonucleic acid

RNA-seq – Ribonucleic acid sequencing

RSV – Respiratory Syncytial Virus

RV – Rhinovirus

SARS-CoV-2 – Severe Acute Respiratory Syndrome Coronavirus 2

SRA – Sequencing Read Archive

SSHRC – Social Sciences Humanities Research Council

TNFSF10 – Trail

TUBB – Tubulin Beta Class I

## Declaration of Academic Achievement

This thesis was written by myself, Dayna Mikkelsen, with revisions and suggestions provided by Dr. Jeremy Hirota. The project management, data interpretation and writing were completed by myself, except for the sections outlined below. Figures were created using BioRender.com.

*Chapter 1:* Leading our team in the Lab2Market program, I recruited interview participants, organized, and conducted the majority of interviews with help from Dr. Jeremy Hirota, Andrew Doxey, Manjot Hunjan, Alex Wu and Victoria Kirkness. Development of VisuFlow was completed by Manjot Hunjan and Alex Wu.

*Chapter 2*: The research presented in this chapter was a collaborative paper, "Validation of CXCL10 as a biomarker of respiratory tract infections detectable by lateral flow immunoassay," which was accepted as a poster to ERS Lung Conference 2022 and will subsequently be submitted for publication. Authorship is shared with Jennifer Aguair. The recruitment of patients, sample collection and processing for the Sunnybrook samples was done by Samira Mubareka and by Jodi Gilchirst, David Bulir, and Marek Smieja for the HRLMP samples. Bioinformatic processing and statistical analysis was completed by Jennifer Aguair and myself. CXCL10 LFA prototype development was completed through an Industry Partnership with Cytodiagnostics by Kha Tram and Catherine Lambert.

*Chapter 3*: Literature research, writing was done by me with revisions and editing by Dr. Jeremy Hirota.

## Impact of COVID Restrictions on the Research Plan

The COVID-19 pandemic severely affected this project, both an advantage and a disadvantage. On the one hand, the samples and funding directly resulted from the COVID-19 pandemic. On the other hand, access to McMaster infrastructure, services and resources was severely limited for most of this project until the final months. Classes and seminars were entirely online. Due to the limited access and training opportunities, Dr. Jeremy Hirota ran laboratory samples, and my training focused on dry lab bioinformatic processes. Managing the collaborative nature of our project, working with an industry partnership, multiple universities, and publicly available datasets posed both a challenge and an opportunity for the completion of this thesis. The thesis, non-traditional as it may be, tapped into entrepreneurial, open-science and collaborative domains. It provided a novel, interdisciplinary and challenging learning opportunity for those involved, attempting to make the most out of the difficult and unforeseen COVID-19 pandemic.

# Chapter 1: Introduction:

## 1.1 The Field of Diagnostics: A Brief History and Looking to the Future

A diagnosis is defined as "the act of identifying a disease from its signs and symptoms."[1] Historically the diagnosis of respiratory diseases and infections relied heavily on clinical clues, such as a cough, fever and wheeze, while today, diagnostics encompass a comprehensive range of functional measurements (e.g. spirometry), advanced imaging techniques (e.g. CT or MRI) and molecular laboratory methods.[2,3] In the context of respiratory tract infections, molecular lab methods are essential tools for identifying the presence of pathogens that can lead to the development or worsening of lung health and disease. Polymerase chain reaction (PCR) methods are clinically useful and implemented widely, allowing for a highly sensitive method to diagnose viral and bacterial pathogens with multiplexed molecular assays[4]. It should be noted that the discovery of PCR in 1985 by Dr. Kary B. Mullis[5] was an advance now widely used and cost-effective but carries limitations of breadth of analysis. For example, PCR is a highly sensitive assay, which means any form of contamination may lead to false positive results.[6] This was observed during the SARS-CoV-2 pandemic, as PCR results could remain positive for months after infection and could not discriminate between active or remnant viral RNA from the original infection.[7] Nucleic acid sequencing has the potential to overcome the limitations of PCR and be widely adopted as costs decrease and infrastructure becomes more readily available. Incorporating advanced molecular technologies for diagnostic tests, including nucleic acid sequencing, into the healthcare system requires a systematic approach involving scientific, economic, and social considerations that challenge the present state to introduce a future state.

> *<u>The present state:</u>*
> *The current workflow to diagnose a respiratory infection varies depending on where a person seeks medical attention and the severity of the illness. Standard practice involves identifying symptoms, a physical examination and past medical history.*[8] *In some cases, diagnostic techniques such as a rapid antigen test, sputum or blood culture, PCR testing and radiological examination may be performed.* [8] *The empirical prescription of antibiotics may be prescribed based on clinical suspicion.*[9] *The determination of the causative agent has a heavy emphasis on the pathogen and rarely integration of biomarkers of host immune response. However, in some viral infections, immunopathology may drive more damage than direct viral injury.*[10] *In addition, some bacterial respiratory pathogens are commensals from the microbiome, making it difficult to distinguish the causative agent. Current practices may lead to the inappropriate prescription of antibiotics, contributing to antimicrobial resistance. In addition, to repeat visits to physicians and suboptimal treatment.*

> *<u>A future state?</u>*
> *I invite you to envision an "ultimate diagnostic test" for respiratory infections. Imagine that from a single self-administered swab taken by an individual at home could be run on a countertop-based device to identify all microorganisms and host metagenomic information. All living microorganisms, including all commensals part of the microbiome, pathogenic organisms, including viruses,*

> *bacteria, and fungi, and associated antimicrobial genes could be identified. In addition, the host transcriptome would be analyzed revealing biomarkers of inflammation and immunopathology. Host biomarkers would provide personalized insights into an individual's response to any pathogens present or perturbation of native microbiome. Additional archived host genome data from prior healthcare interactions could be integrated with host responses and microbial profiling, with single nucleotide polymorphisms or other variants informing host responses. Combining all these data parameters into an algorithm to provide the right treatment(s), at the right time, to the right person. In this future diagnostic state where relevant scientific, economic, and social considerations have been addressed, society experiences increased quality of life and efficient delivery of optimized care. This future can only happen if we believe it can happen.*

In this transition from present state to future state, we must take steps, and all cannot be completed at once. A logical step from lone detection of pathogens is the merging with simultaneous measurement of host responses for the identification of diagnostic or prognostic biomarkers.

In general, host biomarkers may use a variety of biological characteristics, including genetic, transcriptomics or proteomics, to diagnose a disease.[11] Host biomarkers can be used to subtype groups of patients on a mutual underlying biological mechanism and molecular heterogeneity[12] or determine optimal treatment methods.[11] Specifically looking at respiratory infections, host biomarkers offer significant potential benefits for several reasons. Using host biomarkers could help discriminate bacterial and viral infections, aiding in better treatment and reducing the unnecessary prescription of antibiotics.[13] Host biomarkers can also determine disease progression and severity[14] and active and latent tuberculosis infection.[15] Host biomarkers subtyping groups of patients off molecular heterogeneity could lead to the development of companion diagnostic tests to identify patient populations with improved to certain drugs.[16] Lastly, host biomarkers of respiratory infection shed light on the drivers of immunopathology. This could serve as a tool in drug discovery research to help identify the pathways causing the majority of the damage. This opens the possibility of developing host-directed treatment to "ameliorate immunopathology".[17] Collectively, the above real-world examples with clinical utility support a step towards integrating host-biomarker detection for improving the management of respiratory tract infections in a "future state".

### 1.2 Market Assessment for Respiratory Tract Infection Host-biomarker Diagnostics

We concede that a "future state" faces headwinds that are beyond scientific and technical advances and must consider marketplace realities influenced by economics, social norms, and personnel skilled in the relevant arts. Market research is able to inform on these important variables that influence the adoption of innovative technologies. Combining market research training and exploration with laboratory research is an integrated approach to match R&D activities with market pain points. To aid in R&D activities, market research during this thesis included activities within the Lab2Market (L2M) program run by Toronto Metropolitan University (previously Ryerson University) and Mitacs. The program's goal was to explore the

commercial viability of our academic research through market research interviews and conversations with stakeholders in the field. The topic examined throughout the L2M program was "A combination host-pathogen rapid diagnostic for decentralized monitoring of respiratory infections." Over 25 interviews were conducted with key players in the field of clinical diagnostics, including ER physicians, family physicians, diagnostic device manufacturers, scientific consultants, and several others (**Figure 1.1A**). Interview notes from all the discussions were combined in an infographic (**Figure 1.1B**) highlighting the key themes and topics discussed.

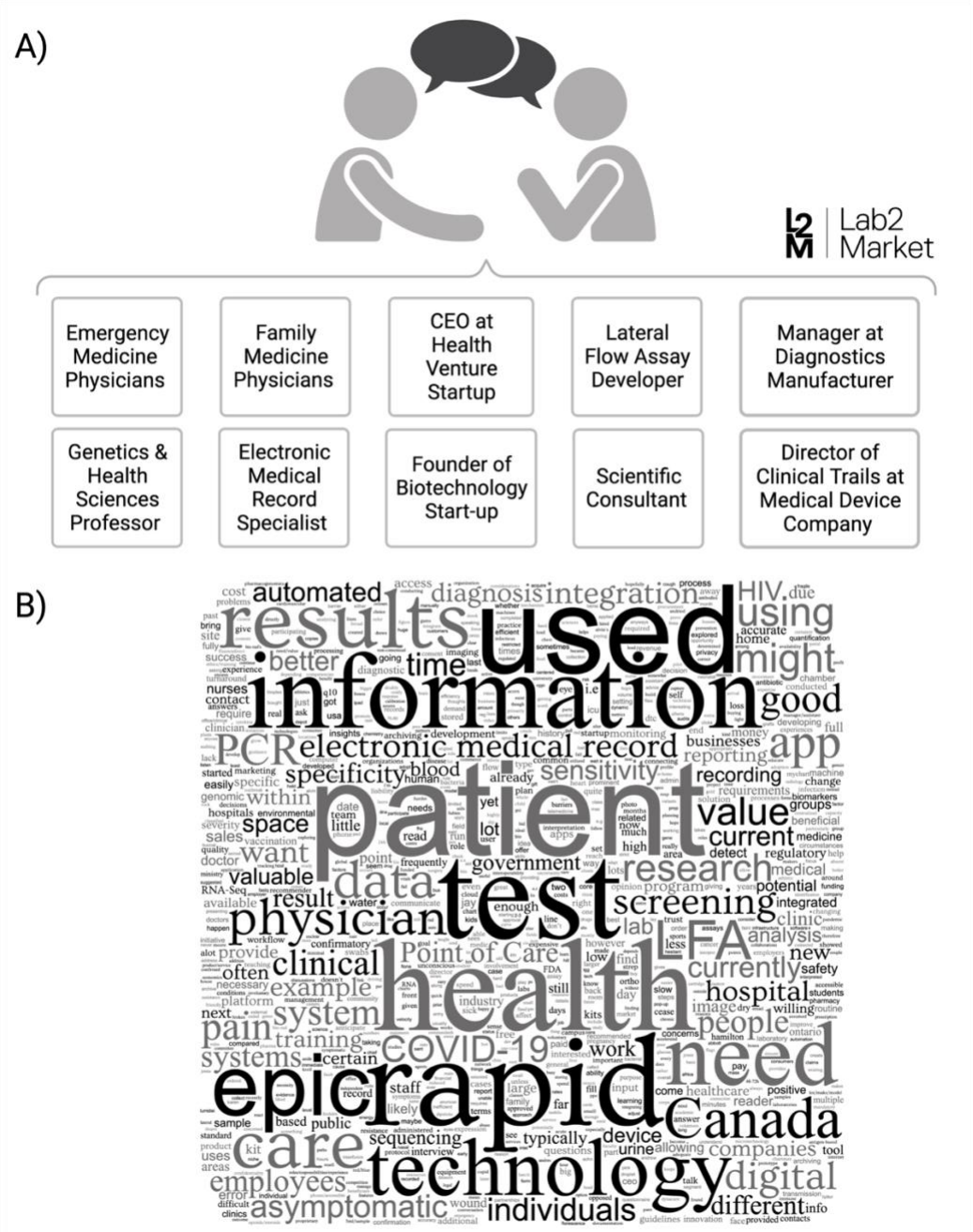



**Figure 1.1 Summary of Lab2Market Program A)** Stakeholders in clinical diagnostics interviewed for qualitative market research through the Lab2Market (L2M) Program. **B**) Word Cloud generated from 25 market research interview notes. Created using https://www.wordclouds.com/

The program's original purpose was to gain a broader understanding of the path to market for host-biomarker based technologies monitoring respiratory tract infections. However, the conversations naturally revealed systemic challenges within Canada's healthcare system and the field of molecular diagnostics. The findings discovered through the L2M program helped provide a solid foundation on diagnostics through hearing first-hand experiences. This guided future work completed throughout this project, emphasizing clinical applications. We were interested in the physician's perspective on the clinical utility of a host biomarker for respiratory infections. While one physician we spoke to thought it was interesting, they highlighted how limited funding is within Canadian hospitals and the importance of considering economic implications and change to existing workflows. They emphasized several issues concerning electronic medical record integration and patient communication. The concept of a combination host-pathogen diagnostic test for respiratory infections was far more complex and specific than the pain points that physicians were facing day to day. Another issue highlighted through these interviews was that a diagnostic providing a "maybe" or suggestive answer is not very useful in the emergency room. They referred to concerns regarding SARS-CoV-2 rapid tests demonstrating that if a host biomarker was integrated into clinical care, it would need to prove clinical utility, a topic that is addressed more comprehensively in the discussion of this thesis.

If host-biomarkers were integrated into clinics, the testing platform would need to integrate with existing diagnostics workflow and infrastructure. One interview highlighted that Canada lacks facilities or equipment to do routine genome (or transcriptome) sequencing for common respiratory infections. The current workflow to diagnose respiratory infections is to use traditional microbial culture methods or run specific tests with select panels for biomarkers of interest, such as PCR. Whole genome sequencing is routinely used for public health surveillance measures in Canada, including for SARs-CoV-2 variant tracking [18] and foodborne pathogens outbreaks such as E. coli O157:H7.[19] In some instances, whole genome sequencing is utilized clinically for rare diseases and cancer. However, whole genome (or transcriptome) sequencing is not routinely used to diagnose acute respiratory infections; it is mainly reserved for research purposes.

The direct costs of routine sequencing would likely be greater than what is currently budgeted for diagnostic workflows. For example, metagenomics next-generation sequencing using respiratory samples can cost anywhere from $500 - $2500 USD per sample.[20] Not to mention the cost of the sequencing platform itself, data analysis and infrastructure to handle the large amounts of data generated from each sample. In addition, the turnaround time can take 1-10 days [20]; when dealing with a critically ill patient, ten days is too long to wait to prescribe antibiotics. For that reason of cost and turnaround time, we explored a decentralized healthcare method via a point-of-care test and pursued the development of a host biomarker lateral flow assay (LFA).

### 1.3 Point-of-care Technologies and the Rapid Adoption of Lateral Flow Assay Technology

An LFA is a paper-based detection technology that can identify the presence of an analyte within a sample.[21] An LFA is typically composed of a strip which contains a sample pad, conjugate release pad, a membrane with antibodies and an adsorbent pad and is occasionally housed within a plastic cartridge.[21] A sample is placed on the sample pad, and through capillary action, the sample travels down the strip where the sample is exposed to conjugate antibodies, which may or

may not bind to the sample depending on if the analyte is in the sample. The sample continues travelling down the trip to the test line, which will display the result, and the control line, which indicates to the user that the liquid successfully moved through the strip.[21] One of the most commonly used LFA tests is the pregnancy test which identifies human chorionic gonadotropin (HCG) in urine (Patent: US4313734A.[22,23] More recently, the COVID-19 pandemic introduced a variety of LFAs detecting SARS-CoV-2 antigens and host immunoglobulins.[24,25] The widespread roll-out has led to rapid tests being used in schools, workplaces, and homes, which has begun to normalize the concept of self-testing at the point of need.

The main advantages of LFAs are that they are easy to use, inexpensive and have a rapid turnaround time. The average LFA test takes anywhere from 5 to 30 minutes to complete and can usually be used without any additional laboratory equipment.[21,26] Together, these qualities make LFA a great point of care tool in healthcare in decentralized diagnostics, which is especially useful in remote locations. As mentioned before, the COVID-19 pandemic increased the use of LFAs as they became a common household name. This was advantageous for their use as people became familiar with how to operate LFAs and global supply chains are in place to develop these technologies at scale.

Through the L2M program, the interviews with clinical diagnostics stakeholders revealed some disadvantages and concerns regarding LFAs, mainly regarding SARS-COV-2 rapid antigen tests. The first concern was regarding the sensitivity of SARS-CoV-2 rapid tests. A systematic review found that SARS-CoV-2 rapid tests, on average, have a sensitivity of 58.1% in asymptomatic people and 72% in symptomatic people.[27] They observed that the sensitivity of the tests increased during the first week of symptoms and in those with a high Ct (cycle threshold) value from the RT-PCR test, which can indicate a higher viral load.[28]

During our conversation with an ER physician, he stated:

> *"Specificity is not good enough for this purpose – Nasopharyngeal PCR swabs are more accurate so advice would be to wait to formal test and not give rapid test / What is the point of a rapid test if a result confirmation is required anyways? -SM".*

Due to the low sensitivity of SARS-CoV-2 rapid tests, confirmatory PCR tests are often required. The argument challenges the clinical utility of SARS-CoV-2 rapid tests in an environment where life-or-death decisions are made. **However, it does not negate the use of this technology for public health surveillance in general settings such as schools and workplaces.** Several jurisdictions within Canada[29] and globally [30] throughout the pandemic employed SARS-CoV-2 asymptomatic detection strategies using rapid antigen LFAs. Within the EU, Liverpool conducted a pilot program to determine whether it prevented SARS-CoV-2 transmission. The results found that the asymptomatic screening program reduced cases by 21% and hospitalizations by 32% during the pilot[31] providing evidence for the utility of SARS-CoV-2 rapid tests in community settings. It also suggests that this approach could be relevant for other respiratory tract infections.

The last disadvantage of an LFA test is that the final result, the test line, is interpreted manually, which is subjective. The subjective interpretation of positive and negative results by the human

eye can dampen sensitivity and specificity, and the documentation of results is not automated or digitized.[32] For example, Figure 1.2 illustrates how often the search term "faint line covid test" was searched on Google from 2019 to 2022. Since rapid test results are self-reported, the interpretation of a "faint line" being positive or negative plays a role in the sensitivity and specificity of the diagnostic test. It is estimated 30.9% of adults have hyperopia, also called farsightedness, where nearby objects appear blurry,[33] furthering the need for an unbiased interpretation of the LFA result.

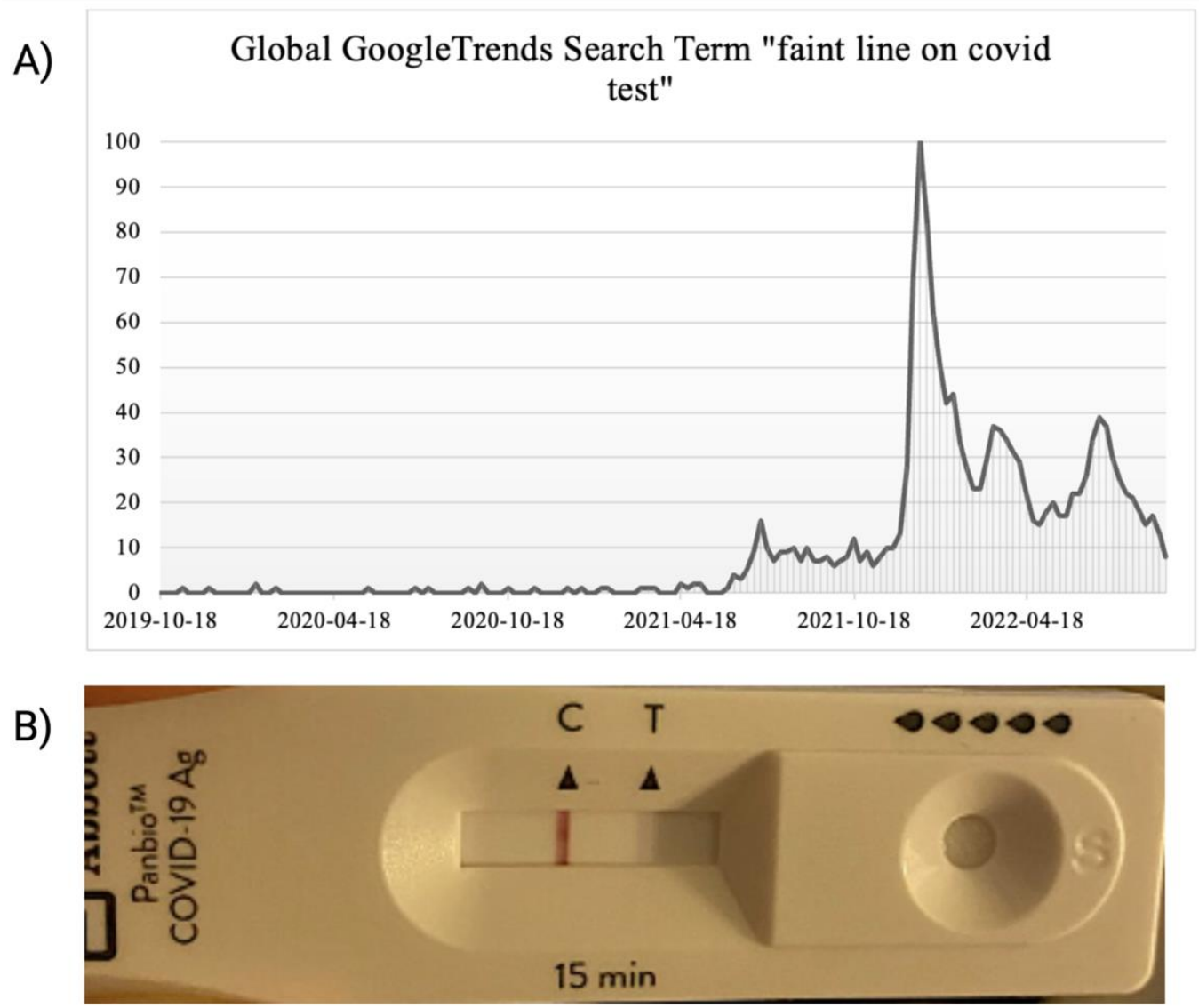



**Figure 1.2 Subjective Interpretation of LFA. A** Google Trends "Interest over time" for the search term "faint line on covid test" from October 2019 to September 2022. Y axis numbers represent Google search interest relative to the highest point on the chart for the given time. Value = 100 is peak popularity for the term. Value = 50 means that the term is half as popular. Value = 0 means not enough data for this term. Data source: Google Trends[34] **B** Example of a faint test line on Abbott Panbio COVID-19 Ag SARS-CoV-2 Rapid test.

Several companies addressed these limitations by developing LFA reader devices that offer an automated readout of the LFA test such as the BD Veritor™ Plus Analyzer.[32,35] The benefits of an LFA readout device are that it offers an objective interpretation, may interpret multiplex tests and quantitative readouts and can provide an automated data management tool integrating with the clinical workflows.[32] Although adding an expense of purchasing a readout device and some require connection to power or a computer, this impedes the low-cost and portable aspect of the LFAs.[32]

In parallel to developing host-biomarker based diagnostics for respiratory tract infections, we addressed this gap by developing VisuFlow©: an open-source cloud framework for smartphone-based LFA diagnostics – improving the interpretation and documentation of results (**Figure 1.3**). It should be noted that the development of VisuFlow© was not the primary focus of the thesis. It attempted to address the multiple pain points regarding LFA interpretation and archiving discovered during market research. Previous work using Samsung SM-P585 tablets and deep learning classified HIV rapid tests and performed better than the standard human eye interpretation by health care staff.[36] This could be very useful in low-middle income countries in which LFA readers are not accessible, but health care staff have access to smartphones or tablets.[36] It is estimated global smartphone adoption in 2021 was 68% and is estimated to increase to 81% by 2025.[37] Smartphone camera technology advancements have continually improved the number of megapixels, allowing for higher quality photos each year.[38] Thus, using existing smartphones as image-capturing devices for LFA interpretation and data management may be helpful in low-resource settings. In contrast, specific features of LFA readout devices, such as uniform lighting conditions and higher image capturing requirements, leave a strong use case for standardized LFA reader devices. Over the course of this thesis, several smartphone LFA reading applications have also been developed such as Abbott's BinaxNOW™ SARS-CoV-2 LFA can be used with their NAVICA smartphone application [39] in addition to existing applications such as BBI Solutions' Novarum™ platform.[40]


**Figure 1.3. VisuFlow©: An open-source cloud framework for smartphone-based LFA diagnostics** – improving the interpretation and documentation of LFA results. Front-end user interface (client end) for smartphone application and validation of the sensitivity/specificity and backend admin dashboard. Development of VisuFlow was completed by Manjot Hunjan and Alex Wu.

Another drawback of most SARS-CoV-2 LFAs is that they do not include any insight into the host response, such as disease outcome and severity. The concept of testing host molecular biomarkers is not entirely novel. Host biomarker-responsive LFAs have historically been used in clinical settings such as the Quidel Triage® D-Dimer Test, which quantifies D-dimer levels for the diagnosis of a pulmonary embolism and/or deep vein thrombosis (DVT) [41] or the pregnancy test mentioned previously.[22,23]

Through the L2M program, the topic of regulation, validation and clinical utility were mentioned frequently. The path to approval by governing regulatory boards such as the Food and Drug Administration (FDA) and Health Canada is a rigorous but vital step in developing a clinical diagnostic test. The diagnostic test must demonstrate analytical validity, clinical validity and clinical utility.[16] Demonstrating analytical validity determines how well a test can detect a specific biomarker. Analytical validity would determine the assay's detection limit, the reproducibility of results, and analytical sensitivity and specificity [16]. The next step is determining the clinical validity, specifying how well the test can predict a clinical diagnosis or disease[16]. Last, the clinical utility is determined. Clinical utility focuses on how the diagnostic can improve the standard of care and allow health care professionals to make more informed

decisions. How can the diagnostic assimilate into the clinical workflow? Social, economic, and ethical considerations are also factored into play during this step. As a result, the development, validation, and regulatory approval of a novel diagnostic test require a lot of time and money. An in-vitro diagnostic device, on average, may cost $34 million and take six years to reach market approval.[42,43]

Canada is in a challenging ecosystem for innovative diagnostics since the health care system is publicly paid, and there is limited funding. This may impede the adoption of novel diagnostics as more pressing demands, such as labour shortages and baseline operations, take precedence. This emphasizes the need to demonstrate clinical benefit clearly; without that, the technology will likely not be adopted into standard care.

In conclusion, the L2M qualitative research taught us a great deal of information. It provided a high-level view of the status of clinical diagnostics in Canada. Host-based diagnostics offer great potential to improve the standard of care for acute respiratory infections by discriminating between viral or bacterial infections, active or latent infections, or prognosis of disease severity. We focused on LFA testing platforms which offer a low-cost point-of-care technology. However, the L2M program highlighted concerns of sensitivity and clinical use of LFA tests with respect to SARS-CoV-2. The conversations revealed more about government regulations and clinical pain points than the genes and proteins involved in my thesis research.

### 1.4 Open Science and the Influence on Academic and Industrial R&D in Respiratory Tract Infections

In 2021, the United Nations Educational, Scientific and Cultural Organization (UNESCO) published a recommendation on Open Science at their General Conference in Paris, defining open science as:

> "*an inclusive construct that combines various movements and practices aiming to make multilingual scientific knowledge openly available, accessible and reusable for everyone, to increase scientific collaborations and sharing of information for the benefits of science and society, and to open the processes of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community."* [44]

Open science is an initiative that includes four main components [45,46]:
**1**) open access to data
**2**) open data sharing
**3**) open-source code and sharing of procedures and methodologies and
**4**) open peer review

Canada's Tri-agency (CIHR, NSERC and SSHRC) has followed suit with UNESCO's recommendations and adopted this policy which requires recipients of their funds to have open access to publications for 12 months after publication.[47] In addition, CIHR has requested that molecular data be deposited into proper repositories when publishing research.[47] In addition, the

US recently committed to requiring all taxpayer-funded work to be open access by 2025, in addition to data from these papers.[48,49]

The field of genomics and -omics research has been a pioneer of open science through open-source data-sharing initiatives for several years, dating back to the Human Genome Project, which was completed in 2003.[50] Since then, there have been numerous open access public repositories, including GenBank ®, NCBI Gene Expression Omnibus (GEO)[51], Sequence Read Archive (SRA), Human Cell Atlas[52], Comprehensive Antibiotic Resistance Database (CARD)[53], Harmonizome[54] and many more.

There are several advantages to open data sharing in research. First, researchers across the globe can reuse data to verify results or conduct further analysis.[48] Scientists can quickly test hypotheses and save resources. Open data sharing in research can also provide larger sample sizes. This is especially useful within the field of artificial intelligence, which requires large datasets to train algorithms.[55]

Some caveats to open science and sharing publicly available datasets are as follows. The EU Horizon Europe requests that shared data be findable, accessible, interoperable, and reusable (FAIR).[46,48] Many researchers have expressed a strong interest in accessing publicly available data.[56] Yet, many do not share data due to a lack of time, limited rights to share data, and inadequate infrastructure.[56] Metadata is only useful if the industry's terms and language are consistent and agreed upon. This has motivated the implementation of "ontologies" when building open science databases which are defined as "systematic representations of knowledge that can be used to integrate and analyze large amounts of heterogeneous data".[57]

These classifications can properly label heterogenous, ad-hoc clinical data into labelled groups, which helps to address inconsistencies when combining large datasets.[57] Another hurdle within open data sharing is data privacy concerns for research participants. While this can be over with informed consent from patients, requesting permission for future consent for further experiments and anonymizing data[58], there have been instances where a patient's anonymity has been compromised and could be re-identified through genomic data.[59] This highlights the importance of controlling access to certain databases and informing research participants of potential risks [59].

The COVID-19 pandemic expedited the use of open data platforms and open access to publications.[60] Open science transitioned from an optional task to a compulsory action for global collaboration. On January 30th, 2020, at the Institut Pasteur, the SARS-Cov-2 genome was released [61] from there, enormous research efforts were undertaken in a wide range of disciplines, including diagnostics and therapeutics, collaborating with worldwide expertise to combat the pandemic. By March 2020, Pfizer announced its partnership with the German company, BioNTech, to co-develop the BNT162b2 (Comirnaty®) mRNA Covid-19 vaccine, which was estimated to have prevented over 110,000 deaths in the USA.[62,63] Examples of open data platforms released during COVID-19 include John Hopkins Corona Virus Resource Center, Our World in Data COVID-19, Nextstrain, GISAID, COVID-19 Genes Cornell and several COVID-19 waste water surveillance platforms.[18,64–69] In addition, COVID-19 allowed for several host genome and transcriptome datasets from blood and nasopharyngeal swabs.[64,70–72] There has been an increase in host-RNA sequencing datasets measuring differential gene expression between

groups with infections and controls from clinical samples.[64,73–75] Combining open science and access to RNA-sequencing data comparing differentially expressed genes (DEGs) would allow large study populations to validate mRNA signatures for specific infections. As stated above, establishing reference standards in RNA-sequencing methods and consistency in metadata labelling is essential to combine findings from multiple host-transcriptomic immune response studies.

In conclusion, open science and data-sharing initiatives are essential to scientific collaboration and innovation. Some view it as a moral obligation to provide Canadian taxpayers with the scientific information generated from tax-paying dollars. However, providing adequate infrastructure to handle data and using consistent terminologies and ontologies are essential to make the shared data useful. In addition, the privacy and consent of research participants must be protected.

### 1.5 Conclusion and Presentation of Thesis Proposal

Integrating these concepts, the field of diagnostics for respiratory infections faces many challenges in the present day, placing a burden on society. Nonetheless, there is plenty of opportunity to drive improvements and advancements through novel research and the advent of next-generation sequencing (NGS). Advances in academic research have demonstrated the capability of characterizing one's microbiome, including pathogenic and commensal organisms and their host transcriptome from one swab.[76] This presents the opportunity to characterize pathogenic viruses, bacteria and fungi alongside the host microbiome and host biomarkers of inflammation and response to therapeutics— moving away from the inappropriate prescription of antibiotics and suboptimal treatment advancing toward the reality of personalized medicine.

The host transcriptome, in particular, is often under looked. At the same time, the emphasis is placed on pathogen detection, but host immunopathology may play a more significant part in viral infections than previously thought. While this all seems compelling, through our market research and conversations with stakeholders within the field, several barriers exist to the clinical implementation of NGS for respiratory infections. First, the cost of NGS far exceeds the public funding in Canada for ARI diagnostics. Second, alternative testing platforms such as LFAs struggle with low sensitivity of LFAs and are not always clinically useful. LFAs on the market do not provide intel into the host immune response.

Open science is a great tool to progress scientific research, saving time and resources. However, this is only useful when there is consistency in data, and it is findable, accessible, interoperable and reusable (FAIR).[46] The COVID-19 pandemic revealed just how vital open science is – having the SARS-CoV-2 genome released in early 2020 and 12 months later, a vaccine was developed, saving over 100,000 lives.

Incorporating these ideas, we hypothesized that host-expressed genes involved in the antiviral response are an effective biomarker of viral infection. Utilizing the breadth of data released during the COVID-19 pandemic through open science, RNA-sequencing and microarray gene expression datasets of viral infection were screened from upper respiratory tract infection samples. The expression patterns of three antiviral genes, *CXCL10*, *CXCL11*, and *TNFSF10*,

were determined for multiple respiratory tract infections: respiratory syncytial virus, rhinovirus, influenza A and SARS-CoV-2 in respiratory mucosa samples. The gene expression patterns suggested that CXCL10 and CXCL11 elevations were consistent for multiple viruses, correlated with higher SARS-CoV-2 viral load, and had a lower variance throughout COVID-19 infection compared to TNFSF10. Mass-spectrometry demonstrated that CXCL10 protein was absent in healthy individuals' saliva samples. SARS-CoV-2 infected individuals indicated significant elevations in CXCL10 protein concentration in saliva. Subsequently, a prototype LFA for CXCL10 protein was developed with a sensitivity of 2ng/ml in human saliva for potential downstream applications in monitoring host-biomarkers in respiratory tract infections.

**1.6 References:**


1. Merriam-Webster. *Merriam-Webster Dictionary: Diagnosis*. Merriam-Webster.com dictionary Accessed October 20, 2022. https://www.merriam-webster.com/dictionary/diagnosis

2. Macfarlane JT, Worboys M. The Changing Management of Acute Bronchitis in Britain, 1940–1970: The Impact of Antibiotics. *Med Hist*. 2008;52(1):47-72.

3. Geddes D. The history of respiratory disease management. *Med Abingdon Engl UK Ed*. 2016;44(6):393-397. doi:10.1016/j.mpmed.2016.03.006

4. Zhang S, Zhang W, Tang YW. Molecular Diagnosis of Viral Respiratory Infections. *Curr Infect Dis Rep*. 2011;13(2):149-158. doi:10.1007/s11908-011-0168-x

5. Mullis KB. The Unusual Origin of the Polymerase Chain Reaction. *Sci Am*. 1990;262(4):56-65.

6. Garibyan L, Avashia N. Research Techniques Made Simple: Polymerase Chain Reaction (PCR). *J Invest Dermatol*. 2013;133(3):e6. doi:10.1038/jid.2013.1

7. Binnicker MJ. Can Testing Predict SARS-CoV-2 Infectivity? The Potential for Certain Methods To Be Surrogates for Replication-Competent Virus. *J Clin Microbiol*. 2021;59(11):e00469-21. doi:10.1128/JCM.00469-21

8. Blasi F, Dimopoulos G. *Textbook of Respiratory and Critical Care Infections*. Jaypee Brothers Medical Publishers; 2014. http://ebookcentral.proquest.com/lib/mcmu/detail.action?docID=4699836

9. Dellit TH, Owens RC, McGowan JE, et al. Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America guidelines for developing an institutional program to enhance antimicrobial stewardship. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2007;44(2):159-177. doi:10.1086/510393

10. Rouse BT, Sehrawat S. Immunity and immunopathology to viruses: what decides the outcome? *Nat Rev Immunol*. 2010;10(7):514-526. doi:10.1038/nri2802

11. Maslove DM, Tang B, Shankar-Hari M, et al. Redefining critical illness. *Nat Med*. 2022;28(6):1141-1148. doi:10.1038/s41591-022-01843-x

12. Neff RA, Wang M, Vatansever S, et al. Molecular subtyping of Alzheimer's disease using RNA sequencing data reveals novel mechanisms and targets. *Sci Adv*. 2021;7(2):eabb5398. doi:10.1126/sciadv.abb5398

13. Ross MH, Zick BL, Tsalik EL. Host-Based Diagnostics for Acute Respiratory Infections. *Clin Ther*. 2019;41(10):1923-1938. doi:10.1016/j.clinthera.2019.06.007

14. Lorè NI, De Lorenzo R, Rancoita PMV, et al. CXCL10 levels at hospital admission predict COVID-19 outcome: hierarchical assessment of 53 putative inflammatory biomarkers in an observational study. *Mol Med*. 2021;27(1):129. doi:10.1186/s10020-021-00390-4

15. Estévez O, Anibarro L, Garet E, et al. An RNA-seq Based Machine Learning Approach Identifies Latent Tuberculosis Patients With an Active Tuberculosis Profile. *Front Immunol*. 2020;11. Accessed September 24, 2022. https://www.frontiersin.org/articles/10.3389/fimmu.2020.01470

16. Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat Rev Genet*. 2016;17(5):257-271. doi:10.1038/nrg.2016.10

17. Wallis RS, O'Garra A, Sher A, Wack A. Host-directed immunotherapy of viral and bacterial infections: past, present and future. *Nat Rev Immunol*. Published online June 7, 2022:1-13. doi:10.1038/s41577-022-00734-z

18. COVID-19 Wastewater Surveillance in Ontario. Public Health Ontario. Accessed September 26, 2022. https://www.publichealthontario.ca/en/Data-and-Analysis/Infectious-Disease/COVID-19-Data-Surveillance/Wastewater

19. Rumore J, Tschetter L, Kearney A, et al. Evaluation of whole-genome sequencing for outbreak detection of Verotoxigenic Escherichia coli O157:H7 from the Canadian perspective. *BMC Genomics*. 2018;19:870. doi:10.1186/s12864-018-5243-3

20. Ramachandran PS, Wilson MR. Metagenomics for neurological infections - expanding our imagination. *Nat Rev Neurol*. 2020;16(10):547-556. doi:10.1038/s41582-020-0374-y

21. Koczula KM, Gallotta A. Lateral flow assays. *Essays Biochem*. 2016;60(1):111-120. doi:10.1042/EBC20150012

22. Kasetsirikul S, Shiddiky MJA, Nguyen NT. Challenges and perspectives in the development of paper-based lateral flow assays. *Microfluid Nanofluidics*. 2020;24(2):17. doi:10.1007/s10404-020-2321-z

23. Vaitukaitis JL, Braunstein GD, Ross GT. A radioimmunoassay which specifically measures human chorionic gonadotropin in the presence of human luteinizing hormone. *Am J Obstet Gynecol*. 1972;113(6):751-758. doi:10.1016/0002-9378(72)90553-4

24. Hsiao WWW, Le TN, Pham DM, et al. Recent Advances in Novel Lateral Flow Technologies for Detection of COVID-19. *Biosensors*. 2021;11(9):295. doi:10.3390/bios11090295

25. Mina MJ, Peto TE, García-Fiñana M, Semple MG, Buchan IE. Clarifying the evidence on SARS-CoV-2 antigen rapid tests in public health responses to COVID-19. *Lancet Lond Engl*. 2021;397(10283):1425-1427. doi:10.1016/S0140-6736(21)00425-6

26. Eltzov E, Guttel S, Low Yuen Kei A, Sinawang PD, Ionescu RE, Marks RS. Lateral Flow Immunoassays – from Paper Strip to Smartphone Technology. *Electroanalysis*. 2015;27(9):2116-2130. doi:10.1002/elan.201500237

27. Dinnes J, Deeks JJ, Berhane S, et al. Rapid, point-of-care antigen tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database Syst Rev*. 2021;2021(3):CD013705. doi:10.1002/14651858.CD013705.pub2

28. Fajnzylber J, Regan J, Coxen K, et al. SARS-CoV-2 viral load is associated with increased disease severity and mortality. *Nat Commun*. 2020;11(1):5493. doi:10.1038/s41467-020-19057-5

29. Rosella LC, Agrawal A, Gans J, Goldfarb A, Sennik S, Stein J. Large-scale implementation of rapid antigen testing system for COVID-19 in workplaces. *Sci Adv*. 2022;8(8):eabm3608. doi:10.1126/sciadv.abm3608

30. Rapid antigen testing (RAT). Ministry of Health NZ. Accessed October 15, 2022. https://www.health.govt.nz/covid-19-novel-coronavirus/covid-19-health-advice-public/covid-19-testing/rapid-antigen-testing-rat

31. Univeristy of Liverpool. *Liverpool Covid-SMART Community Testing Pilot: Evaluation Report*. University of Liverpool, Department of Health and Social Care (DHSC).; 2021:66. Accessed October 20, 2022. https://www.liverpool.ac.uk/media/livacuk/research/Mass,testing,evaluation.pdf

32. Faulstich K, Gruler R, Eberhard M, Lentzsch D, Haberstroh K. Handheld and Portable Reader Devices for Lateral Flow Immunoassays. In: Wong R, Tse H, eds. *Lateral Flow Immunoassay*. Humana Press; 2009:1-27. doi:10.1007/978-1-59745-240-3_9

33. Hashemi H, Fotouhi A, Yekta A, Pakzad R, Ostadimoghaddam H, Khabazkhoob M. Global and regional estimates of prevalence of refractive errors: Systematic review and meta-analysis. *J Curr Ophthalmol*. 2017;30(1):3-22. doi:10.1016/j.joco.2017.08.009

34. GoogleTrends. Google Trends: faint line on covid test. Google Trends. Published October 18, 2019. Accessed October 20, 2022. https://trends.google.com/trends/explore?date=2019-10-20%202022-09-20&q=faint%20line%20on%20covid%20test

35. BD. How the BD Veritor™ Plus System Works. BD Veritor. Accessed October 19, 2022. https://bdveritor.bd.com/en-us/main/bd-veritor-plus-system/how-it-works

36. Turbé V, Herbst C, Mngomezulu T, et al. Deep learning of HIV field-based rapid tests. *Nat Med*. 2021;27(7):1165-1170. doi:10.1038/s41591-021-01384-9

37. GSMA | The Mobile Economy - The Mobile Economy. Accessed December 7, 2021. https://www.gsma.com/mobileeconomy/

38. Wood CS, Thomas MR, Budd J, et al. Taking connected mobile-health diagnostics of infectious diseases to the field. *Nature*. 2019;566(7745):467-474. doi:10.1038/s41586-019-0956-2

39. Navica App & BinaxNOW™ COVID-19 Ag test card. Accessed October 19, 2022. https://www.globalpointofcare.abbott/en/product-details/navica-binaxnow-covid-19-us.html

40. Novarum™ | Smartphone Diagnostics for Rapid Testing - BBI Solutions | Lateral Flow, Reagents and POCT. Accessed October 19, 2022. https://www.bbisolutions.com/en/digital/what-is-novarum

41. Linkins LA, Lapner ST. Review of D-dimer testing: Good, Bad, and Ugly. *Int J Lab Hematol*. 2017;39:98-103. doi:10.1111/ijlh.12665

42. Chin CD, Linder V, Sia SK. Commercialization of microfluidic point-of-care diagnostic devices. *Lab Chip*. 2012;12(12):2118. doi:10.1039/c2lc21204h

43. Drlik M. How Much Does it Cost to Develop a Medical Device? FOUR METHODS TO ESTIMATE MEDICAL DEVICE AND POINT OF CARE DIAGNOSTIC DEVELOPMENT COSTS.

44. UNESCO Recommendation on Open Science - UNESCO Digital Library. Accessed September 27, 2022. https://unesdoc.unesco.org/ark:/48223/pf0000379949

45. Ciriminna R, Scurria A, Gangadhar S, Chandha S, Pagliaro M. Reaping the benefits of open science in scholarly communication. *Heliyon*. 2021;7(12). doi:10.1016/j.heliyon.2021.e08638

46. Wilkinson MD, Dumontier M, Aalbersberg IjJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3(1):160018. doi:10.1038/sdata.2016.18

47. Government of Canada I. *Tri-Agency Open Access Policy on Publications*. Government of Canada; 2016. Accessed October 20, 2022. https://science.gc.ca/site/science/en/interagency-research-funding/policies-and-guidelines/open-access/tri-agency-open-access-policy-publications

48. Musen MA. Without appropriate metadata, data-sharing mandates are pointless. *Nature*. 2022;609(7926):222-222. doi:10.1038/d41586-022-02820-7

49. Brainard J, Kaiser J. U.S. to require free access to papers on all research it funds. *Science*. 2022;377(6610):1026-1027. doi:10.1126/science.ade6577

50. International Consortium Completes Human Genome Project. Genome.gov. Accessed September 26, 2022. https://www.genome.gov/11006929/2003-release-international-consortium-completes-hgp

51. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res*. 2013;41(Database issue):D991-995. doi:10.1093/nar/gks1193

52. Regev A, Teichmann SA, Lander ES, et al. The Human Cell Atlas. *eLife*. 6:e27041. doi:10.7554/eLife.27041

53. Alcock BP, Raphenya AR, Lau TTY, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res*. 2020;48(D1):D517-D525. doi:10.1093/nar/gkz935

54. Rouillard AD, Gundersen GW, Fernandez NF, et al. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database*. 2016;2016(baw100). doi:10.1093/database/baw100

55. Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: A call for open science. *Patterns*. 2021;2(10). doi:10.1016/j.patter.2021.100347

56. Tenopir C, Rice NM, Allard S, et al. Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLOS ONE*. 2020;15(3):e0229003. doi:10.1371/journal.pone.0229003

57. Haendel MA, Chute CG, Robinson PN. Classification, Ontology, and Precision Medicine. *N Engl J Med*. 2018;379(15):1452-1462. doi:10.1056/NEJMra1615014

58. Choudhury S, Fishman JR, McGowan ML, Juengst ET. Big data, open science and the brain: lessons learned from genomics. *Front Hum Neurosci*. 2014;8. Accessed September 14, 2022. https://www.frontiersin.org/articles/10.3389/fnhum.2014.00239

59. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying Personal Genomes by Surname Inference. *Science*. 2013;339(6117):321-324. doi:10.1126/science.1229566

60. Besançon L, Peiffer-Smadja N, Segalas C, et al. Open science saves lives: lessons from the COVID-19 pandemic. *BMC Med Res Methodol*. 2021;21(1):117. doi:10.1186/s12874-021-01304-y

61. Institut Pasteur. SARS-CoV-2, COVID-19 epidemic: the Institut Pasteur's response, scientific research and discoveries in 2020. Institut Pasteur. Published February 16, 2021. Accessed September 19, 2022. https://www.pasteur.fr/en/research-journal/reports/sars-cov-2-covid-19-epidemic-institut-pasteur-s-response-scientific-research-and-discoveries-2020

62. Di Fusco M, Marczell K, Deger KA, et al. Public health impact of the Pfizer-BioNTech COVID-19 vaccine (BNT162b2) in the first year of rollout in the United States. *J Med Econ*. 2022;25(1):605-617. doi:10.1080/13696998.2022.2071427

63. Lamb YN. BNT162b2 mRNA COVID-19 Vaccine: First Approval. *Drugs*. 2021;81(4):495-501. doi:10.1007/s40265-021-01480-7

64. Butler D, Mozsary C, Meydan C, et al. Shotgun transcriptome, spatial omics, and isothermal profiling of SARS-CoV-2 infection reveals unique host responses, viral diversification, and drug interactions. *Nat Commun*. 2021;12(1):1660. doi:10.1038/s41467-021-21361-7

65. CDC. National Wastewater Surveillance System. Centers for Disease Control and Prevention. Published May 16, 2022. Accessed September 26, 2022. http://www.cdc.gov/healthywater/surveillance/wastewater-surveillance/wastewater-surveillance.html

66. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*. 2020;20(5):533-534. doi:10.1016/S1473-3099(20)30120-1

67. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. Kelso J, ed. *Bioinformatics*. 2018;34(23):4121-4123. doi:10.1093/bioinformatics/bty407

68. Khare S, Gurry C, Freitas L, et al. GISAID's Role in Pandemic Response. *China CDC Wkly*. 2021;3(49):1049-1051. doi:10.46234/ccdcw2021.255

69. Ritchie H, Mathieu E, Rodés-Guirao L, et al. Coronavirus Pandemic (COVID-19). *Our World Data*. Published online March 5, 2020. Accessed September 26, 2022. https://ourworldindata.org/coronavirus

70. Lieberman NAP, Peddu V, Xie H, et al. In vivo antiviral host transcriptional response to SARS-CoV-2 by viral load, sex, and age. Cadwell K, ed. *PLOS Biol*. 2020;18(9):e3000849. doi:10.1371/journal.pbio.3000849

71. Ng DL, Granados AC, Santos YA, et al. A diagnostic host response biosignature for COVID-19 from RNA profiling of nasal swabs and blood. *Sci Adv*. 2021;7(6). doi:10.1126/sciadv.abe5984

72. Genome Canada. CANADIAN COVID-19 GENOMICS NETWORK: Generating accessible and usable genomics data to inform policy and public health decisions. Published online March 2022.

73. Langelier C, Kalantar KL, Moazed F, et al. Integrating host response and unbiased microbe detection for lower respiratory tract infection diagnosis in critically ill adults. *Proc Natl Acad Sci U S A*. 2018;115(52):E12353-E12362. doi:10.1073/pnas.1809700115

74. Oved K, Cohen A, Boico O, et al. A Novel Host-Proteome Signature for Distinguishing between Acute Bacterial and Viral Infections. *PLOS ONE*. 2015;10(3):e0120012. doi:10.1371/journal.pone.0120012

75. Ross M, Henao R, Burke TW, et al. A comparison of host response strategies to distinguish bacterial and viral infection. *PLoS ONE*. 2021;16(12):e0261385. doi:10.1371/journal.pone.0261385

76. Rajagopala SV, Bakhoum NG, Pakala SB, et al. Metatranscriptomics to characterize respiratory virome, microbiome, and host response directly from clinical samples. *Cell Rep Methods*. 2021;1(6):100091. doi:10.1016/j.crmeth.2021.100091

# Chapter 2: Thesis Manuscript

## Characterization of CXCL10 as a Biomarker of Respiratory Tract Infections Detectable by Open-source Lateral Flow Immunoassay

Dayna Mikkelsen[1,2]*, Jennifer A. Aguiar[3]*, Julia Danieli[3], Prakriti Chhabra[3], Benjamin J-M Tremblay[4], Manjot S. Hunjan[3], Ulrich Eckhard[5], Jodi Gilchrist[6], David Bulir[6], Marek Smieja[6], Samira Mubareka[7], Catherine Lambert[8], Kha Tram[8], Andrew C. Doxey[1,2,3], and Jeremy A. Hirota[1,2,3,9]

1 Department of Medicine, McMaster University, Hamilton, ON, Canada
2 Firestone Institute for Respiratory Health – Division of Respirology, Department of Medicine, McMaster University, Hamilton, ON, Canada
3 Department of Biology, University of Waterloo, Waterloo, Ontario, Canada
4 Centre for Research in Agricultural Genomics, Carrer de la Vall Moronta, Cerdanyola del Vallès, Barcelona, Catalonia, Spain.
5 Department of Structural Biology, Molecular Biology Institute of Barcelona, CSIC, Barcelona Science Park, Baldiri Reixac, Barcelona, Catalonia, Spain.
6 Hamilton Regional Laboratory Medicine Program, Hamilton, ON, Canada
7 Sunnybrook Research Institute, Toronto, ON, Canada
8 Cytodiagnostics Incorporated, Burlington, ON, Canada
9 Division of Respiratory Medicine, Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada, V6H 3Z6

* these two authors contributed equally

**Corresponding Author:**
Jeremy Hirota, PhD
Firestone Institute for Respiratory Health – Division of Respirology, Department of Medicine, McMaster University, Hamilton, ON, L8N 4A6
Phone: 905-518-0745
Email: hirotaja@mcmaster.ca

**Acknowledgements:** The authors acknowledge all of the authors of the publicly available datasets used who have facilitated analysis through email correspondence. We would also like to acknowledge Reina Ditta, Amanda Hodge, and Guillaume Pare from the Genomics and Molecular Epidemiology Lab at McMaster University for their core services for microarray gene expression assays.

**Funding:** This work was supported by the Roche Canada COVID Open Innovation Challenge, Ontario Government COVID-19 Rapid Research Fund, and the Thistledown Foundation/FastGrants.org. Hirota JA is supported by the Canada Research Chairs program and an Ontario Early Researcher Award. Doxey AC is supported by NSERC and an Ontario Early Researcher Award.

## 2.1 Introduction

Prior to COVID-19, lower respiratory infections ranked as the 4th leading cause of death responsible for ~ 2.6 million deaths annually.[1] The majority of these deaths occurred in children under five years old and those over the age of 70.[1] The COVID-19 pandemic has increased the global burden of disease attributable to respiratory infections, with > 6 million fatalities as of July 2022 [2] with age functioning as a significant risk factor that has spared children from the greatest morbidity and mortality.[3]

The morbidity and mortality associated with respiratory tract infections are dominant aspects that contribute to a global burden calculation, but these measures must also be considered in conjunction with the economic costs.[4] Economic costs of respiratory infections can be broken down into direct and indirect costs. Direct costs of respiratory infections are incurred through interactions with the healthcare system and may include physician visits and medication costs. Indirect costs may include absenteeism for patients and caregivers.[5] Total costs of pneumonia and acute lower respiratory infections exceed $25 billion annually in the European Union for a population of ~450 million people.[6] Extrapolating these costs on a global population level estimates ~$400 billion in annual costs across all jurisdictions. When including COVID-19, these costs are much higher. On a global scale, the economic burden of COVID-19 led to a -3.1 % decrease (equal to -$2.4 trillion USD) in the annual global GDP[7], with total estimates potentially as high as $16 trillion USD.[8] The COVID-19 pandemic demonstrated that in dire circumstances, respiratory infections can cause unprecedented global economic recession and massive disruptions to global supply chains.[9] While the economy has begun to recover from COVID-19, the risk of novel emerging infectious diseases and SARS-CoV-2 variants of concern remains high. The significant direct and indirect costs of respiratory infections warrant the development of a diverse set of approaches that can help mitigate economic burden.

Diversity in respiratory tract virus infections exists with differential impacts on morbidity, mortality, and economic burden. Across common respiratory tract viral infections, the young, elderly, individuals with chronic respiratory diseases, and pregnant mothers have elevated risks for morbidity and mortality.[10] Common viral respiratory tract pathogens include rhinoviruses (RV), respiratory syncytial virus (RSV), influenza (FluA and FluB), and coronaviruses.[3,11] Depending on the pathogen, respiratory viral infections may vary in presentation from a mild illness restricted to upper airways that is self-resolving (RV)[12] to acute respiratory distress syndrome that requires ventilation and is associated with fatalities (SARS-CoV-2).[3] Globally it has been estimated that 2% of all deaths related to respiratory dysfunction are associated with influenza causing approximately ~400 000 deaths every year.[13] RV cause relatively mild illness in the majority of the population; although they contribute to a substantial number of medical visits and absent days of work every year as increased morbidity is observed in immunocompromised individuals.[14] RV infections contribute to a substantial number of acute asthma exacerbations in the community.[15] RSV is common viral infection in children under 5 years old that is associated with hospitalizations and wheeze.[16,17] In a longitudinal study design of 92 children,[18] all but one individual had an RSV infection prior to age 2, suggesting that a large proportion of children have the potential to be infected with RSV. In 2015, there were approximately 33 million RSV infections globally which resulted in three million hospitalizations and just under 60,000 deaths in children younger than 5 years old.[19] Both RV

and RSV infections pose as a risk factor for subsequent development of wheeze and childhood asthma.[20,21] Understanding core mechanisms common to respiratory tract viral pathogenesis and host-responses to infections may provide biomarkers for at-risk patient populations that guide interventions aimed at reducing morbidity, mortality, and economic costs.

To combat respiratory tract viral infections, the respiratory mucosa participates in a wide range of innate and adaptive immune responses to control and eliminate pathogens. Innate immune responses of the respiratory mucosa are initiated by germ-line encoded pattern recognition receptors (PRRs) that bind to molecular motifs that are associated with extrinsic pathogen associated molecular patterns (PAMPS) or intrinsic damage associated molecular patterns (DAMPS). PRR recognition of non-self PAMPS or self-DAMPS initiates intracellular signaling cascades that upregulate host-defence mechanisms defined by temporally specific gene expression patterns.[22] Although these processes are broadly applicable to all respiratory tract viral infections, diversity in host responses exist and are influenced by the infectious pathogen and its ability to subvert host immune responses in attempt to replicate and spread.[23,24] Core-host innate immune responses that are shared between respiratory tract viral infections may provide diagnostic and prognostic potential beyond simple pathogen detection alone.[25,26] Importantly, host innate immune responses are not merely beneficial and can be attributed to pulmonary damage via uncontrolled by inflammation and immunopathology.[27]

The dominant innate immune response induced upon respiratory tract viral infection and downstream of PRR signaling is the induction of interferon stimulated genes (ISGs) through Type I and Type III interferons.[24] When a virus enters the respiratory system, viral PAMPs are recognized by the hosts' PRRs, and this triggers the production of interferons (type I and III) and actives the interferon antiviral response. Type I IFNs act in a paracrine and endocrine fashion through the Type I IFN receptor (IFNAR) while Type III IFNs signal through IFNLR1.[24] Downstream signaling from these receptors merges on JAK family of kinases and STAT family of transcription factors, leading to transcriptional activation.[28] The interferon antiviral response induces a number of different ISGs that generate protein products that inhibit distinct parts of viral life cycles. [24,28] Aspects that the interferon antiviral response may coordinate include cellular defences to virus binding, internalization, replication, packaging, and release. Importantly, in conjunction with ISG products produced *within* the cell to combat viral life cycle, other ISG products are released *outside* the cell to signal to non-infected cells in the local and systemic environment. Secreted ISG protein products that are also chemokines may provide early biomarker signals that are prognostic for respiratory tract viral infections. Three of these chemokines include TNFSF10 (TRAIL) and the CXC chemokines: CXCL10 and CXCL11 which are all induced by interferon. Previous research has found these three proteins CXCL10/CXCL11/TNFSF10 to be induced by several respiratory viral infections both *in-vitro* and *in-vivo* experiments.[29–31]

CXCL10, also called interferon-γ inducible protein 10 kDa (IP-10), is a secreted ISG and plays a key role in fighting viral infections.[32(p10)] This chemokine has been shown to have altered mRNA expression levels in RV, RSV, and SARS-CoV-2 infections.[33–35] In addition, serum samples revealed high levels of CXCL10 protein in H5N1 influenza infection and SARS-CoV-2.[36,37] Recently, Lorè et al[37] showed CXCL10 serum levels were a robust predictor of COVID-19 outcomes, with higher levels being related to increased mortality. CXCL11, also known as C-X-

C motif chemokine ligand 11, is another chemokine which shares the same CXCR3 receptor as CXCL10 and is also secreted.[38] Both CXCL10 and CXCL11 are induced by interferon-γ and play a role in recruitment of T cells and natural killer (NK) cells as part of the immune response.[38]. CXCL11 has shown a similar trend in up-regulated in SARS-CoV-2 infections *in-vitro* [39] and *in-vivo.*[35,40] Tumour necrosis factor-related apoptosis-inducing ligand (TRAIL), from the gene *TNFSF10,* is a secreted protein which can induce apoptosis and plays a role in the immune response to destroy virally infected cells.[41] *TNFSF10* has been found to be up regulated during viral infection.[42] Together, these three genes *CXCL10, CXCL11* and *TNFSF10* are cytokines that are also stimulated by the IFN response and may be useful biomarkers for respiratory tract viral infections.

Respiratory tract infections caused by viruses have sought a massive burden on health care system for years – although the recent COVID-19 pandemic has exacerbated this exponentially. COVID has ushered in a new acceptance of rapid diagnostics that includes lateral flow assays (LFAs) being used outside hospitals such as schools and workplaces. LFAs are useful for testing directly on-site rather than operating through a centralized testing laboratory. The advantages of LFA tests are that they are quick, inexpensive, and easy to use. Although, due to lower sensitivity and specificity, they are typically preferred as a screening method and require a confirmation test such as PCR.[43] SARS-CoV-2 rapid tests such as the Abbott Pan Bio detect the SARS-CoV-2 pathogen but do not provide any insight into the individual's host response and disease trajectory. However, Roche's Elecsys® IL-6 test uses electrochemiluminescence immunoassay to quantify IL-6 levels in blood with potential applications for identifying severe disease in COVID-19 patients.[44,45] Leveraging LFA technology for detection of biomarkers prognostic of respiratory tract infection may provide opportunities to intervene selectively and aggressively in those most at risk of poor outcomes.

ISGs play a major role in the antiviral defense and measuring extracellular protein products of ISGs could be potential biomarkers of infection. CXCL10, CXCL11 and TNFSF10 are involved in the antiviral immune response during a variety of respiratory infections. However, the feasibility and performance of these ISGs as useful and measurable clinical biomarkers from an easily accessible sample format remains unknown. In the present study, we had the overarching objective of defining the expression patterns of the secreted ISG products CXCL10, CXCL11, and TNFSF10 for multiple respiratory tract infections including RSV, RV, influenza A and SARS-CoV-2 in respiratory mucosa samples to inform the development of a host-biomarker point of care LFA tool. Gene expression levels from upper airway samples suggested that CXCL10 and CXCL11 elevations were consistent across multiple viruses, correlated with higher SARS-CoV-2 viral load and had the lower variance over course of COVID-19 infection compared to TNFSF10. Deep proteomic profiling using mass-spectrometry revealed CXCL10 was absent in oral samples from healthy individuals. CXCL10 levels were measured from the saliva of SARS-CoV-2 infected individuals and showed significant elevations in CXCL10 protein concentration. We proceeded to develop a prototype LFA for CXCL10 protein with a sensitivity of 2ng/ml in human saliva. Our work provides a foundation for further exploration of CXCL10 as a host biomarker relevant in respiratory tract viral infections with potential diagnostic and prognostic value.

## 2.2 Materials and Methods

**2.2.1 Human ethics:** Procurement of nasopharyngeal swab (NPS) and saliva samples from consented study subjects was approved by Hamilton integrated Research Ethics Board (HiREB 4914T and 5099T).

**2.2.2 Data Resources:** A combination of publicly available gene and protein datasets (Table 1) and in-house generated gene datasets from Hamilton Regional Laboratory Medicine Program and Sunnybrook Hospital were used for present study.

**Table 2.1** Publicly Available Data Resources Accessed for Analysis

| Data Resource and Citation | Corresponding Figure | Sample Type | Gene Expression Method | Data Composition and Notes |
|---|---|---|---|---|
| Harmonizome - https://maayanlab.cloud/Harmonizome/ Citation Link: https://pubmed.ncbi.nlm.nih.gov/27374120/ | Figure 2.1 | Various | Multiple microarray platforms | 366 sets of genes differentially expressed following virus perturbations from the GEO Signatures of Differentially Expressed Genes for Viral Infections dataset.<br><br>mRNA expression profiles in cell lines or tissues: Calu-3 cells, HAE cultures, see supplementary for all. |
| Host Gene Expression in Nasal and Blood Samples for the Diagnosis of Viral Respiratory Infection<br>Citation Link: https://pubmed.ncbi.nlm.nih.gov/30339221/<br><br>GSE117827 | Figure 2.2 | Nasal mid-turbinate swab | Microarray-Affymetrix Human Clariom-D Chips | Total n = 24<br>n = 6 Control<br>n = 6 Respiratory Syncytial Virus<br>n = 12 Rhinovirus<br>Control = children (3mo-18years) having ambulatory surgery for nonacute conditions<br>Cases = children (3mo-18years) hospitalized for acute respiratory illness positive with positive results only for a single virus using BioFire FilmArray Respiratory Panel |
| In vivo antiviral host response to SARS-CoV-2 by viral load, sex, and age.<br>Citation Link: https://pubmed.ncbi.nlm.nih.gov/32898168/ GSE152075 | Figure 2.2 | Nasopharyngeal swabs | RNA-sequencing | Total n = 484. Subjects with suspected SARS-CoV-2 infection underwent an nasopharyngeal swab based RT-PCR test for the SARS-CoV2 N1 gene<br>n = 54 SARS-CoV-2 PCR test negative<br>n = 430 SARS-CoV-2 PCR test positive |
| Upper airway gene expression differentiates COVID-19 from other acute respiratory illnesses and reveals suppression of innate immune responses by SARS-CoV-2<br>Citation Link: https://pubmed.ncbi.nlm.nih.gov/33203890/ GSE156063 | Figure 2.2 | Nasopharyngeal swabs | RNA-sequencing | Total n = 234. Subjects were under investigation for suspected SARS-CoV-2 infection.<br>n = 100 non-viral acute respiratory illnesses (ARIs)<br>n = 41 SARS-CoV-2 PCR test negative<br>n = 93 SARS-CoV-2 PCR test positive |
| Shotgun transcriptome, spatial omics, and isothermal profiling of SARS-CoV-2 infection reveals unique host responses, viral diversification, and drug interactions<br>Citation Link: https://pubmed.ncbi.nlm.nih.gov/33712587/ and https://covidgenes.weill.cornell.edu/ dbGaP Study Accession: phs002258.v1.p1 | Figure 2.3 | Nasopharyngeal swabs | RNA-sequencing | Total n = 735. Subjects were under investigation for suspected SARS-CoV-2 infection.<br>n = 519 SARS-CoV-2 PCR test negative<br>n = 216 SARS-CoV-2 PCR test positive |
| Ultra-deep and quantitative saliva proteome reveals dynamics of the oral microbiome<br>Citation Link: https://pubmed.ncbi.nlm.nih.gov/27102203/ | Figure 2.5 | Oral swabs | Mass spectrometry-based proteomics | Total n = 8<br>n = 4 female<br>n = 4 male<br><br>All subjects were healthy, non-smoking healthy individuals aged 24 to 40 years with Caucasian backgrounds and assessed at two times during the day (morning and evening). |

### 2.2.3 Overview of Publicly Available Accessed Datasets:

*In-vitro* cell culture experiments*: In-vitro*, pre-processed transcriptomic datasets were examined using the Harmonizome resource[46] (https://maayanlab.cloud/Harmonizome/. From Harmonizome, the dataset labelled "GEO Signatures of Differentially Expressed Genes for Viral Infections" was accessed for hypothesis testing for elevations in *CXCL10*, *CXCL11*, and *TNFSF10* following viral exposure *in vitro* from a variety of experimental conditions. The "GEO Signatures of Differentially Expressed Genes for Viral Infections" dataset contained 366 individual datasets of mRNA expression profiles using microarray technology for cell lines or tissues after viral infection. These were filtered down to 199 microarray datasets as all non-human experiments and non-respiratory viruses were excluded. The remaining 199 datasets came from 17 GSE studies including the following viruses: influenza A, human coronavirus, SARS-CoV like virus isolated from bats, a mouse-adapted SARS-coronavirus, SARS-CoV mutant strain that does not express the accessory protein open reading frame 6 (delta ORF6), human metapneumovirus, wild-type infectious clone-derived SARS-CoV, measles, RSV and RV, with time points ranging from 0- 96 hours post inoculation. *In-vitro* cell systems included but were not limited to Calu-3, HAE, and A549 cell lines (Supplementary Table 1). From the 199 microarray datasets, the studies in which *CXCL10, CXCL11* and *TNFSF10* were in the top 300 or bottom 300 differentially expressed were counted. The log fold change and number of datasets were plotted using Graph Pad Prism. This Harmonizome method of data retrieval is a conservative representation of the number of datasets where the candidate genes are differentially regulated as they must be in the top or bottom 300 log fold changed genes. Three representative housekeeping genes *ACTB, TUBB* and *GADPH* were analyzed in parallel as a frame of reference.

Upper Airway Sampling: Four datasets of upper airway swabs (NPS or nasal mid-turbinate) samples were curated to determine *CXCL10, CXCL11* and *TNFSF10* gene expression under different respiratory viral infections. Host transcriptomic data on RV and RSV infections were from pediatric populations between 3 months and 18 years admitted to hospital for acute respiratory illness and had confirmed PCR detection of pathogen.[47] All included subjects had not received any immunosuppressants for 30 days or antibiotics for 7 days. Control subjects were of same age but present in the hospital setting for ambulatory surgery for non-acute conditions. Microarray gene expression analysis was performed on the mid-turbinate nasal swab with data retrieved from GSE117827. A total of n = 6 RSV positive (2 males, 4 females), n = 12 RV positive (8 males and 4 females) and n = 6 negative control (6 males) samples were analyzed with mean age 31 months +/- 39 months (standard deviation).
Host transcriptomic data on SARS-CoV-2 infections were from adult populations from three studies that each had distinct subject enrolment strategies and demographics available. In the first dataset, a total of n = 430 SARS-CoV-2 positive (176 males. 201 females, 53 unknown) with median age 54 years (range 2-98) and n = 54 negative control (30 males, 24 females) median age 46.5 years (range 12-90) year samples were analyzed from GSE152075.[35] For the second dataset, a total of n = 93 SARS-CoV-2 positive (no other pathogenic respiratory virus) (50 females, 43 unknown) with mean age 47 years, n = 41 other viral ARI (19 females, 22 unknown) mean age 51 years and 100 non-viral ARI (55 females, 45 unknown) mean age 57 years samples were analyzed from GSE156063.[48] For the third dataset, SARS-CoV-2 status was determined by RT-PCR and other respiratory virus were detected by metagenomic next generation sequencing. The

non-viral ARI group included bacterial RTIs and other non-infectious respiratory diseases. A total of n = 216 SARS-CoV-2 and n = 519 negative samples were analyzed from Genotypes and Phenotypes dbGAP (accession #38851 and ID phs002258.v1.p1).[40] To determine the association between positive SARS-CoV-2 status and *CXCL10, CXCL11* and *TNFSF10* expression, GSE152075 and GSE156063 were accessed. To determine the association between SARS-CoV-2 viral load and *CXCL10, CXCL11* and *TNFSF10* expression a web portal for an RNA-sequencing dataset was accessed (https://covidgenes.weill.cornell.edu/). We extracted the log2 fold change and q values for *CXCL10, CXCL11* and *TNFSF10* for positive vs negative comparison group and "Viral Level Continuous" comparison group using SVA correction and included all reads. "Viral Level Continuous" comparison group converted SARS-CoV-2 qRT-PCR cycle threshold (Ct) values into a continuous variable by converting Ct values where Ct of 15 is equal 1.0 and a Ct over 40 is taken as 0. The goal of this was to determine if *CXCL10, CXC11* or *TNFSF10* gene expression correlated with SARS-CoV-2 viral load.
Oral Sampling: To determine the expression level of CXCL10, CXCL11, and TNFSF10 at the protein level in healthy individuals, an ultra-deep liquid chromatography-mass spectrometry-based analysis of the human saliva proteome was accessed that included 8 subjects (4 males and 4 females) (PXD 003028).[49] All subjects were healthy and asymptomatic with no oral inflammation, pathologies, or prescribed medications. Each subject was sampled twice, once in the morning and once after breakfast. An aliquot of the morning sample underwent an additional fractionation step for deeper analysis.

### 2.2.4 Overview of Analysis of In-House Datasets:

Upper Airway Sampling: Three datasets of NPS upper airway samples were generated to determine *CXCL10, CXCL11* and *TNFSF10* gene expression under influenza A and SARS-CoV-2 infection. For influenza A infections, NPS samples from patients suspected of respiratory tract infection were collected in the first 6 months of 2020 through the Hamilton Regional Laboratory Medicine Program (HRLMP) during COVID-19 screening. NPS swabs (Copan, Italia) were collected and stored in universal transport media prior to multiplex PCR analysis for a panel of respiratory tract infections. A total of n = 8 influenza A positive (5 males, 3 females) and n = 14 negative control (7 males, 7 females) samples were analyzed with mean age 78.2 +/- 11.9 (standard deviation). For comparison of fatal and non-fatal SARS-CoV-2 infections, NPS samples from positive for COVID-19 were collected in the first 6 months of 2020 through the Sunnybrook Hospital and Research Institute during regional COVID-19 testing programs. NPS swabs (Copan, Italia) were collected and stored in universal transport media prior to RNA isolation. A total of n = 6 fatal (5 males, 1 female) and n = 19 non-fatal (9 males, 10 females) samples were analyzed with mean age 62.4 +/- 21.6 (standard deviation). For determining the variability of *CXCL10, CXCL11*, and *TNFSF10* over time following a positive SARS-CoV-2 test, PCR confirmed COVID-19 positive patients underwent repeated NPS sampling followed by storage in McMaster Molecular Media and RNA isolation. A total of n = 6 subjects (5 females, 1 male – 72.8 years mean age +/-13.3 years) were analyzed over the period in which each subject was an inpatient (range from 12 to 26 days). All RNA isolation was performed using New England Biolabs Monarch Total RNA Miniprep Kit according to manufacturer directions. RNA concentration and integrity number was assessed by Agilent Bioanalzyer according to manufacturer directions. Gene expression profiling was performed using the Human Clariom D microarray assay (Applied Biosystems).

Oral sampling: To determine the suitability for measuring CXCL10 protein in saliva as a surrogate for *CXCL10* gene in upper airway samples, we assessed CXCL10 protein by multiplex cytokine array (Eve Technologies, Calgary, Alberta) in COVID-19 inpatients processed by the Hamilton Regional Laboratory Medicine Program. Neat saliva samples were provided from SARS-CoV-2 positive inpatients (n = 3) and SARS-CoV-2 negative healthy controls (n = 6). Samples were spun down at 1000g for 5 minutes to pellet the mucus in samples with a supernatant isolated for downstream analysis and kept frozen at -80°C until use. Saliva cytokine levels were quantified using Human Cytokine Array / Chemokine Array 71-plex (Eve Technologies, Calgary, Alberta, Canada).

### 2.2.5 Bioinformatic Processing and Statistical Analysis:

For microarray datasets, raw intensity values and annotation data were downloaded. Probe definition files for the Clariom D human microarray chip were retrieved from Bioconductor and probes were annotated with Ensembl IDs in R (version 4.1.0). All gene expression data from a given study were unified into a single data matrix that was then normalized by robust multiarray average (RMA) normalization. Normalized expression levels for *CXCL10, CXCL11,* and *TNFSF10* were extracted.

For RNA-seq datasets, raw count values and annotation data were downloaded directly from the corresponding author's GitHub. Counts were normalized using the limma package in R (version 3.14) with the design matrix including virus, sex, and age annotation data. Normalized expression levels for *CXCL10, CXCL11,* and *TNFSF10* were extracted. Differential expression analysis was performed using the DESeq2 package in R (version 3.14).
Gene expression levels were tested for significant differences between viral infections or other phenotypic data of interest via pairwise Student's t-tests with Benjamini–Hochberg multiple testing correction or via ANOVA followed by Tukey Honest Significant Difference *post-hoc* tests using the stats R package (version 3.6.2). Gene expression box plots were generated with GraphPad Prism 9 (GraphPad Software Inc., USA) and heat maps were generated using pheatmap R package (version 1.0.12). For heat maps, normalized expression levels are scaled by gene. Gene expression box plots were expressed as mean and standard errors of the mean (SEM) with unpaired t-tests performed to compare the means of two groups. Where three groups were compared, a one-way ANOVA was performed with a Bonferroni correction for multiple comparisons. To determine variance in *CXCL10*, *CXCL11*, and *TNFSF10* gene expression over time with SARS-CoV-2 infection, the host gene of interest was expressed as percent change from first period of inpatient sampling (set as time=0) for each individual patient. Line graphs and box plots were generated with GraphPad Prism 9 (GraphPad Software Inc., USA). Differences were considered statistically significant when $p < 0.05$.
To assess protein levels in healthy human saliva, quantitative proteomic data corresponding to ProteomeXchange dataset PXD003028 [49] were downloaded from the MaxQB Database.[50] Reverse hits, common contaminants including keratins, and low confidence hits were removed, and data consistency verified using Pearson's correlation analysis within the Perseus platform.[51] Label free quantification intensities were subsequently summed to give a robust read-out for average protein levels for the 5551 proteins quantified. More than 95% proteins were identified in at least 3 samples, and nearly 80% across all 8 samples.

### 2.2.6 CXCL10 LFA Prototype Development:

Antibody pair selection process: Commercially available antibodies and reagents were selected for assay development to ensure reproducibility and robustness between labs. Antibody pairs for the detection of CXCL10 were identified from commercially available sources that were amenable to ELISA applications. A recombinant anti-human CXCL10 mouse IgG-monoclonal (MAB2661), anti-mouse IgG goat-polyclonal (AF-266-NA), and recombinant CXCL0 protein (Product 266-IP) were selected from R&D Systems (Toronto, Ontario, Canada).

Conjugation of monoclonal CXCL10 antibody to 40nm gold: Conjugation was performed by using an N-Hydroxysuccinimide (NHS) coupling kit from Cytodiagnostics following manufacturer directions (Product CGN5K-40-2, Burlington, Ontario, Canada). Briefly, the monoclonal CXCL10 antibody was rehydrated in the supplied Protein Resuspension Buffer to a final concentration of 0.5 mg/mL. A single conjugation reaction was performed by the addition of the supplied Reaction Buffer and CXCL10 solution to the lyophilized NHS-activated gold nanoparticle pellet. The reaction was gently mixed for 1 hr before adding 10 µL of the supplied Quencher Solution followed by 10 µL of 10% bovine serum albumin. The 40 nm gold conjugate was pelleted by centrifugation (1300x g); the supernatant was removed, and 1 mL of Conjugate Buffer was added to resuspend and wash the conjugate before re-centrifugation to pellet the product. After removing the supernatant, the product was resuspended, and the OD adjusted to 10 using the provided Conjugate Buffer.

Preparation of lateral flow dipsticks: A control line solution (150uL) was prepared using 0.2 mg/mL of the anti-mouse IgG antibody reconstituted in a capture buffer (10 mM HEPES, pH 7.4, 0.1% BSA, 0.5% Methanol). The test line solution (150uL) was prepared with polyclonal CXCL10 antibody at 0.1 mg/mL in capture buffer. Hi-FlowTM Plus 135 (Catalog #HF135MC100, 60mm x 301 mm) nitrocellulose membranes were purchased from EMD Millipore Corporation, Burlington Massachusetts. The test and control solutions were striped in parallel across the non-pre-treated nitrocellulose membranes using a syringe pump set to 2.2mL/min connected Claremont Bio's Automated Lateral Flow Reagent Dispenser set to 3V. The resulting membranes were allowed to fully dry at room temperature for 18 hours before assembly with the wicking pad (Cellulose fibre sample pad, product #CFSP223000, purchased from EMD Millipore; Billerica, Massachusetts, USA) and cut into 4 mm wide strips by a CM5000 Guillotine Cutter. Batches of approximately 150 completed LF strips were made and stored at room temperature for up to 6 months without loss of sensitivity (data not shown).

Lateral flow sensitivity testing: Solutions for LFA sensitivity testing were prepared with 90% (by volume) lateral flow buffer (10mM HEPES, 150 mM NaCl, 0.1% Tween-20, 1% BSA, and 0.5% PEG 8000) and 10% (1 OD final concentration) monoclonal anti-human CXCL10 antibody 40nm gold conjugate. Standards of known concentration were generated by adding recombinant CXCL10 to Reconstitution Buffer (1X PBS pH 7.4, 0.1% bovine serum albumin) to generate 50, 20, 10, 5, and 2 ng/mL final concentrations along with a 0 ng/mL negative control.
Solutions for LFA testing with artificial saliva (product #1700-0316, ASTM E2721-16 with Mucin, pH 7.0; Pickering Laboratories, Mountain View, California, USA) were prepared with 45% (by volume) lateral flow buffer and 45% artificial saliva. Monoclonal anti-human CXCL10 antibody 40nm gold conjugate was added to a final concentration of 1 OD (or 10% by volume).

Standards of known concentration were generated by adding recombinant CXCL10 to Reconstitution Buffer (1X PBS pH 7.4, 0.1% BSA) to generate 400, 50, 20, 10 and 5 ng/mL final concentrations along with a 0 ng/mL negative control. For sensitivity studies in ideal buffer or artificial saliva, five unique replicates of standards were generated with 50ul of the solutions added to a 96-well plate. The solution was allowed to flow completely through the dipstick for 15 min and imaged with band intensity quantification performed using an IUL IPeak lateral flow strip reader (Product # 100033000; Barcelona, Spain) programmed to identify test and control bands.

Testing of LFA with human saliva from healthy subjects: Saliva was collected from healthy subjects with no symptoms of upper or lower respiratory tract infections. Subjects were asked to wash their mouth with water and discard into a sink, followed by 5 minutes of collection of 5 ml of collection (whichever came first). Samples were analyzed as is without centrifugation to represent a real-world point of care setting. Aliquots from 7 individuals were generated to create neat samples and CXCL10 spiked samples (10 ng/ml). LFAs were completed by adding 50ul of sample to a 96 well plate followed by insertion of the dipstick for 15 minutes with qualitative visualization and photo capture completed.

## 2.3 Results

To first investigate the potential association between *CXCL10*, *CXCL11*, and *TNFSF10* with viral infection, pre-processed transcriptomic datasets were examined using the Harmonizome resource (https://maayanlab.cloud/Harmonizome/)[46]. 366 processed microarray datasets under the category "GEO Signatures of Differentially Expressed Genes for Viral Infections"[52,53] were curated and filtered to include only datasets from experiments that used respiratory viruses and in-vitro human samples (**Figure 2.1A**). This filtering reduced the total number of GSE studies to 17 and corresponding microarray datasets to 199. From the 17 different studies, two were from a single time-point and the remaining 15 of the studies were time-series data. A variety of respiratory tract viral pathogens were used in the datasets including influenza A, SARS-CoV variants, human metapneumovirus, measles, RSV and RV (**See Supplementary Table 1**). The Harmonizome database extracted 600 differentially expressed genes following viral infection for each of the 199 microarray datasets which included the top 300 log2 fold change genes with increased expression and bottom 300 log2 fold genes with decreased expression. Of the 199 datasets, there was significantly more datasets with upregulation of *CXCL10*, *CXCL11*, and *TNFSF10* gene transcript relative to datasets where these genes were downregulated. *CXCL10* was found to be significantly up-regulated in 39 datasets and down-regulated in 1(**Figure 2.1B)**. *CXCL11* was up-regulated in 33 datasets. *TNFSF10* was up-regulated in 36 datasets and down-regulated in 3. Among the three genes, *CXCL10* was up regulated in the greatest number of viral infection datasets. The common housekeeping genes *GAPDH*, *TUBB* and *ACTB* were analyzed in all 199 datasets to provide context for the observed upregulation of *CXCL10*, *CXCL11*, and *TNFSF10* (**Figure 2.1C**). The three housekeeping genes showed altered expression in fewer datasets than for *CXCL10*, *CXCL11*, and *TNFSF10*. These results provide evidence that *CXCL10*, *CXCL11*, and *TNFSF10* are up-regulated agnostic to human cell type and pathogen type when filtered for human respiratory tract viral infections and provided a rationale to explore the candidates further with *in-vivo* samples of patients with clinically diagnosed respiratory viral infections.

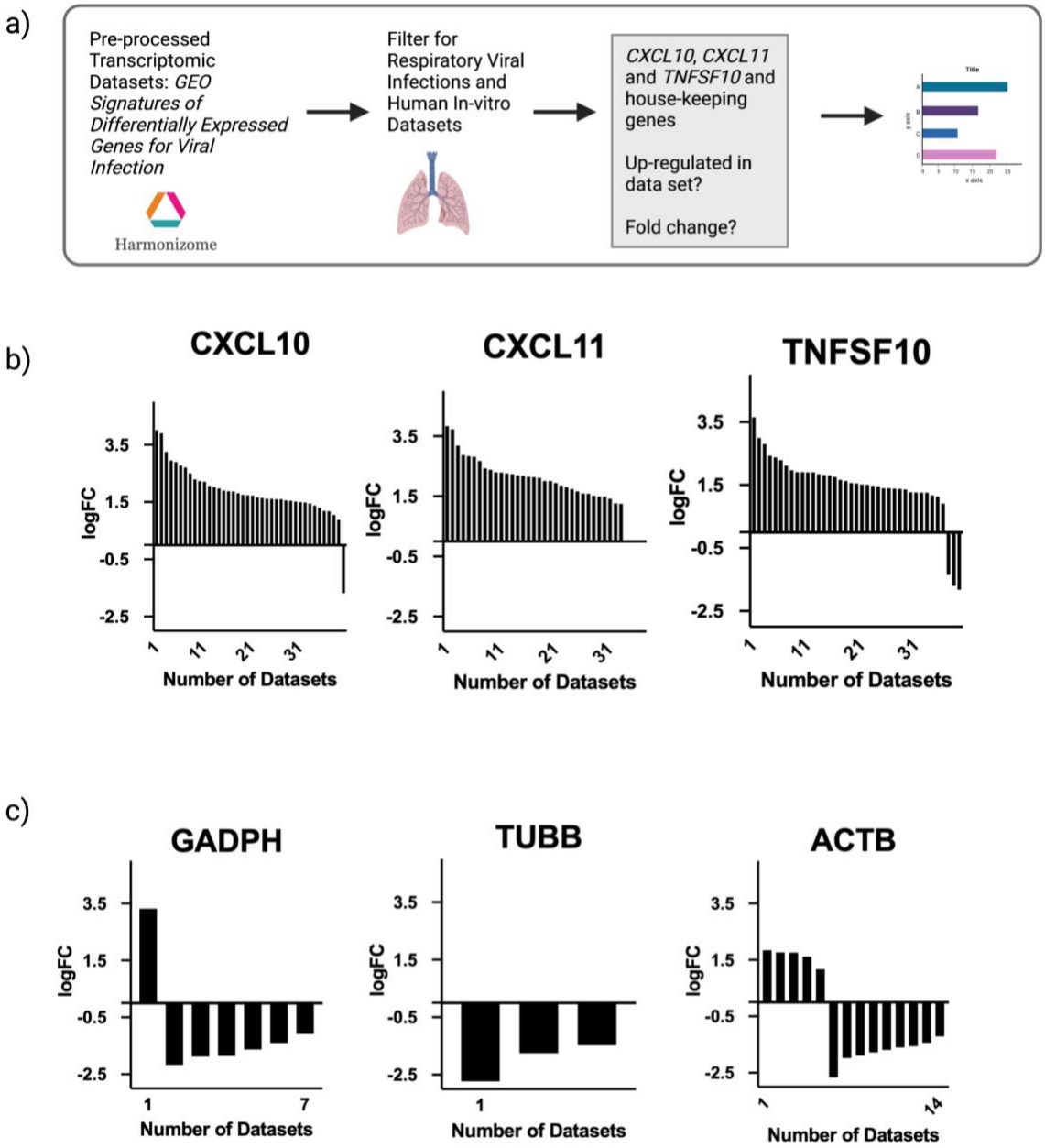



**Figure 2.1: *CXCL10*, *CXCL11* and *TNFSF10* show increased expression *in-vitro* after viral infection. A:** Pre-processed transcriptomic datasets using the Harmonizome (Rouillard et al[46]) database of processed microarray datasets under the category "GEO Signatures of Differentially Expressed Genes for Viral Infections" (Edgar et al[52]; Barrett et al[53]) was accessed for hypothesis testing elevations in *CXCL10*, *CXCL11*, and *TNFSF10* following viral exposure *in vitro* from a variety of experimental conditions. Dataset contained 366 individual datasets of mRNA expression profiles using microarray technology. All non-human experiments and non-respiratory viruses were excluded which filtered down to 199 microarray datasets. The studies in which gene appeared to be at top 300 or bottom 300 differentially expressed were counted. Datasets included but not limited to Calu-3 cell lines, HAE cultures infected with respiratory viruses including but not limited to FluA, SARS-CoV and Human metapneumovirus. **B:** *CXCL10* was found to be up-regulated in 39 independent viral infection datasets, and down-regulated in one. *CXCL11* was up regulated in 33 viral infection datasets. *TNFSF10* was up-regulated in 36 and down-regulated in 3. **C:** Housekeeping genes *GAPDH*, *TUBB* and *ACTB* were analyzed in all 199 and showed different expression levels in fewer datasets than for *CXCL10*, *CXCL11*, and *TNFSF10*. *GADPH* was found to be up-regulated in 1 independent viral infection datasets, and down-regulated in 6. *TUBB* was down in 3 viral infection datasets. *ACTB* was up-regulated in 5 and down-regulated in 9.

To interrogate the expression of *CXCL10*, *CXCL11*, and *TNFSF10* in clinical samples, we analyzed a combination of publicly available and in-house datasets from upper airway samples taken during studies of respiratory tract infections that included RSV, RV, influenza A and SARS-CoV-2 with corresponding non-infected negative controls (**Figure 2.2A**). *CXCL10* and *CXCL11* were up-regulated with RSV infection relative to non-infected controls (**Figure 2.2B** – orange bars, $p < 0.05$) and to samples from RV infected subjects ($p < 0.05$). *CXCL10*, *CXCL11* and *TNFSF10* were not different between NPS samples collected from RV infected subjects relative to non-infected controls (**Figure 2.2B** – blue bars, $p > 0.05$). No significant difference was found between NPS samples from influenza A infected individuals relative to non-infected controls for *CXCL10*, *CXCL11* and *TNFSF10* (**Figure 2.2C**, $p > 0.05$). With the COVID-19 pandemic, a variety of studies from SARS-CoV-2 infected subjects and non-infected controls have become publicly available. Probing a dataset by Lieberman et al[35], we observed increases in *CXCL10*, *CXCL11* and *TNFSF10* in upper airway samples from SARS-CoV-2 infected subjects compared to non-infected negative controls (**Figure 2.2D**, $p < 0.001$). Similarly, analyzing data from Mick et al[48], we observed increases in *CXCL10*, *CXCL11* and *TNFSF10* in samples from SARS-CoV-2 infected subjects compared to non-infected negative controls (**Figure 2.2E**, $p < 0.001$). Collectively, these data from multiple respiratory tract viral infections suggest that *CXCL10*, *CXCL11* and *TNFSF10* positively correlate with infection status, with a robust and validated increase observed during SARS-CoV-2 infection.

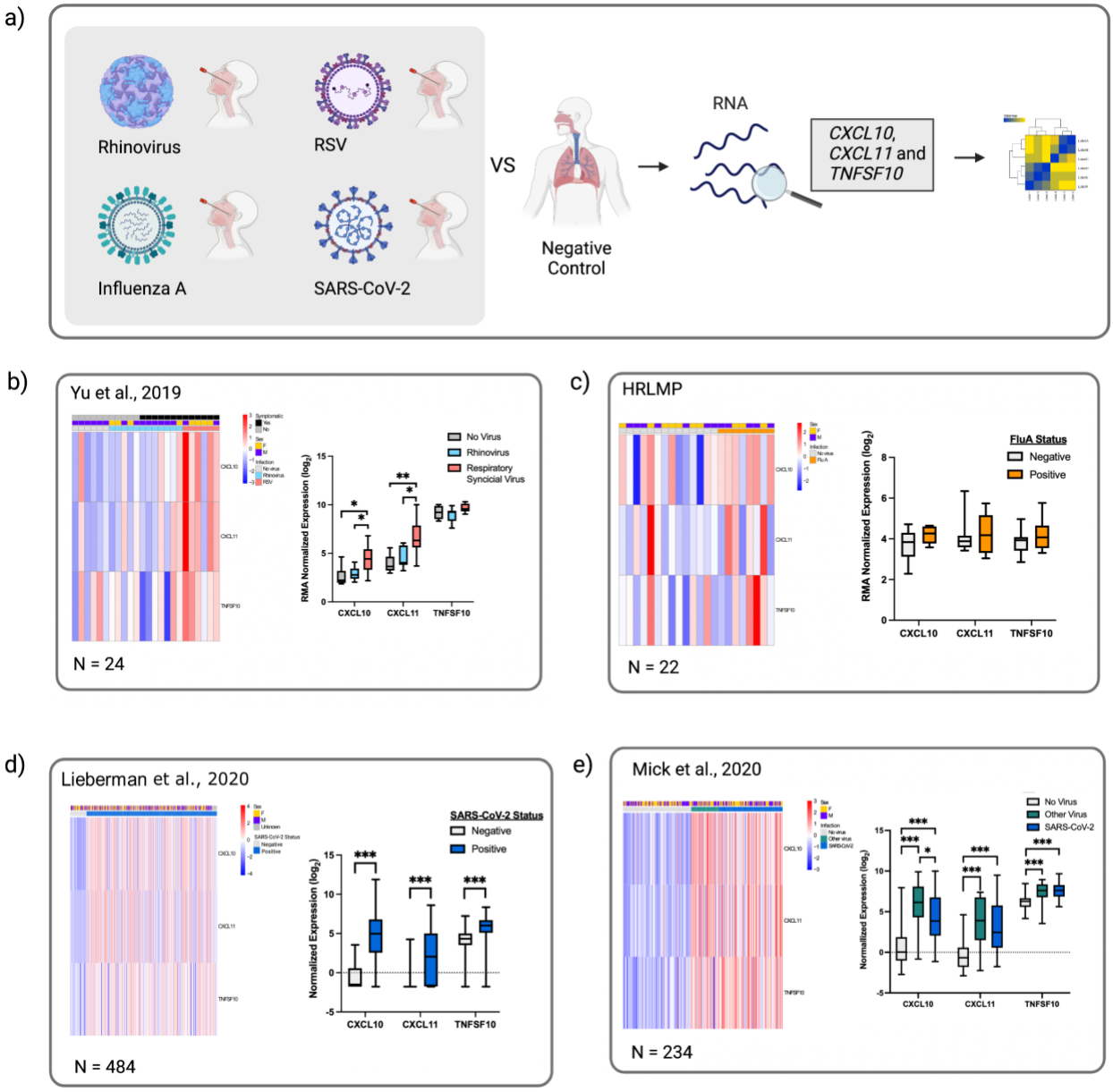



**Figure 2.2. *CXCL10, CXCL11* and *TNFSF10* gene expression in nasopharyngeal swab and nasal mid-turbinate samples of positive vs negative collected from subjects with different viral respiratory tract infections including SARS-Cov-2, FluA, Respiratory Syncytial Virus (RSV), Rhinovirus (RV). B:** From Yu et al[47] GSE117827: *CXCL10*, *CXCL11* and *TNFSF10 gene expression* compared from the mid-turbinate nasal swab of pediatric subjects with RSV, RV and negative controls. Clustered heatmap of log2 expression levels annotated by symptomatic, sex and infection with blue representing decreased expression and red increased expression. On the right, boxplot of RMA normalized expression (log2) (n=24). No significant difference when comparing RV infected NPS to healthy control gene expression for *CXCL10*, *CXCL11* and *TNFSF10* ($p > 0.05$). *CXCL10* and *CXCL11* up-regulation was positively correlated with RSV when compared to control (p=0.016, p=0.006). Comparing Respiratory Syncytial Virus (RSV) to Rhinovirus (RV) infected NPS, *CXCL10* and *CXCL11* were significantly up-regulated in the RSV group (p=0.02, p=0.02). **C:** Hamilton Regional Laboratory Medicine Program (HRLMP) microarray data: *CXCL10*, *CXCL11* and *TNFSF10 gene expression* was compared from NPS of subjects with influenza A (FluA) and negative controls. Clustered heatmap of log2 expression levels annotated by sex and infection status with blue representing decreased expression and red increased expression. On the right, boxplot of RMA normalized expression (log2) (n=22). No significant difference between *CXCL10*, *CXCL11* and *TNFSF10* expression in FluA infection compared to negative control ($p > 0.05$). **D:** From Lieberman et al[35] GSE152075: *CXCL10*, *CXCL11* and *TNFSF10 gene expression* was compared from the NPS of individuals with suspected SARS-CoV-2 infection. Clustered heatmap of log2 expression levels annotated by sex and infection status with blue representing decreased expression and red increased expression. On the right, boxplot of RMA normalized expression (log2) (n=484). *CXCL10*, *CXCL11* and *TNFSF10* expression significantly up-regulated SARS-CoV-2 infection compared to those who tested negative ($p<0.001$). **E:** From Mick et al[48]-GSE156063: *CXCL10*, *CXCL11* and *TNFSF10 gene expression* was compared from the NPS of subjects SARS-CoV-2 positive, SARS-CoV-2 negative but positive for another respiratory virus and no respiratory virus detected by metagenomic next generation sequencing (i.e. non-viral ARI such as bacterial infection). Clustered heatmap of log2 expression levels annotated by sex and infection status with blue representing decreased expression and red increased expression. On the right, boxplot of RMA normalized expression (log2) (n=234). *CXCL10*, *CXCL11* and *TNFSF10* expression was significantly up-regulated SARS-Cov-2 infection compared to healthy control ($p<0.001$). * = $p < 0.05$, ** $p< 0.01$ and *** = $p < 0.001$.

The observation that elevations in *CXCL10*, *CXCL11* and *TNFSF10* transcripts from upper airway swab samples strongly associated with COVID-19 infection, prompted further exploration with additional SARS-CoV-2 study samples with data features of viral load and mortality (**Figure 2.3A**). Using a dataset from 735 subjects (519 PCR confirmed SARS-CoV-2 negative and 216 PCR confirmed SARS-CoV-2 positive) that included metadata on Ct cycle for indication of viral load, we observed that *CXCL10*, *CXCL11* and *TNFSF10* gene expression showed trends for increasing with greater SARS-CoV-2 viral load quantified by PCR Ct value (**Figure 2.3B**). Pooling all SARS-CoV-2 positive samples together and comparing to SARS-CoV-2 negative samples, *CXCL10* showed an upregulation of log2 fold change of 3.4 (q value = 7.38E-31), *CXCL11* showed an upregulation of 2.9 (q value = 4.43E-22) and *TNFSF10* showed an upregulation of 1.0 (q value = 9.39E-23). Conversion of RT-PCR cycle threshold (Ct) values into a continuous variable showed highly significant correlations for the three candidate biomarkers (*CXCL10* - q value = 1.23E-54, *CXCL11* - q value =5.17 E-47, and *TNFSF10* - q value = 4.26E-38). Individuals who tested positive for other respiratory viral infections, under subclass "Other viral infection" appeared to also correlate with higher expression of *CXCL10*, *CXCL11* and *TNFSF10* (**Figure 2.3B –** pink bar for metadata on virus subclass), although no metadata on the other infection status was available.

As viral load may not be an accurate indicator of infection severity in COVID-19 [54,55], we quantified *CXCL10*, *CXCL11* and *TNFSF10* transcript levels in fatal COVID-19 inpatients via microarray gene chip. Using a dataset of 25 subjects (19 PCR confirmed SARS-CoV-2 positive survivors and 6 PCR confirmed SARS-CoV-2 positive deaths) the mean RMA gene expression value for fatal vs non-fatal COVID cases was 4.3 (SD 1.8) vs 3.5 (SD 1.11) for *CXCL10*, 3.7 (SD 0.4) vs 3.3 (SD 0.4) for *CXCL11* and 6.0 (SD 1.5) vs 6.3 (SD 0.8) for *TNFSF10*. In this limited sample size, mean values were higher in COVID-19 fatal cases for *CXCL10* and *CXCL11* but were not significantly different (**Figure 2.3C**, $p > 0.05$). No trends were observed for *TNFSF10*.

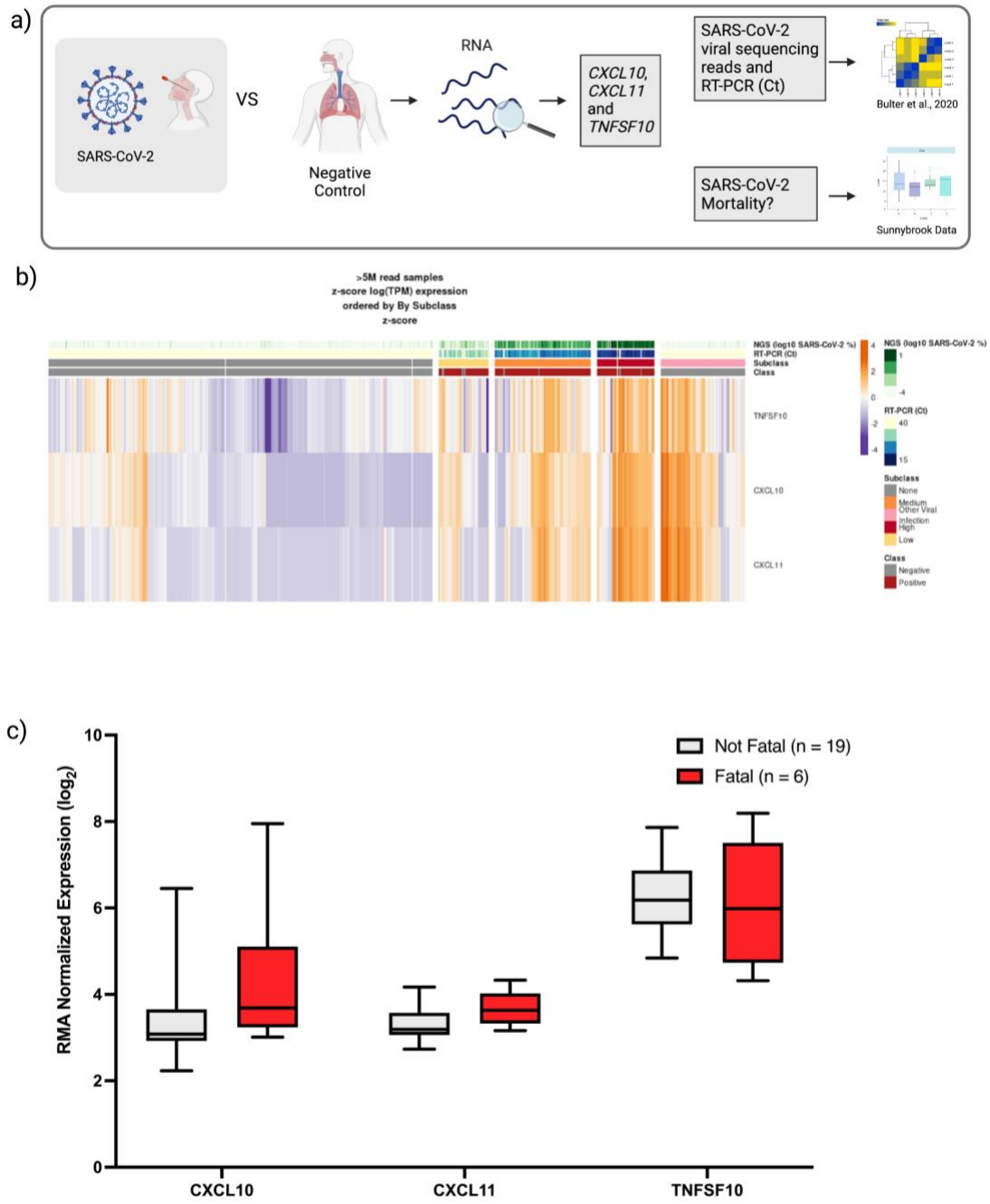



**Figure 2.3: The CXCL10/CXCL11/TNFSF10 gene signature in nasopharyngeal swab samples is positively correlated with SARS-CoV-2 Viral Load.** A: SARS-CoV-2 viral sequencing reads and qPCR cycle thresholds correlate with the CXCL10/CXCL11/TNFSF10 gene signature. (N=735). The "Viral Level Continuous" comparison group converted qRT-PCR cycle threshold (Ct) values into a continuous variable by inverting CT values where Ct = 15 is equal to 1.0 and a Ct > 40 is 0, CXCL10 showed an upregulation of log2 fold change of 6.2 (q value = 1.23E-54), CXCL11 showed an upregulation of log2 fold change of 6.0 (q value =5.17 E-47) and TNFSF10 showed an upregulation of log2 fold change of 1.9 (q value = 4.26E-38). Data and Figure from Butler et al. 2021 - For research purposes only. All rights reserved. © Mason Lab and Weill Cornell Medicine, 2020, Cem Meydan @ Mason Lab). B: Mortality of COVID-19 patients is associated with an elevated CXCL10/CXCL11/TNFSF10 gene signature at the time of original patient sampling. No significant correlation was found ($p > 0.05$).

A limitation of leveraging swab samples at the time of diagnosis and examining host responses is that all samples are not collected at the same time during the course of infection. As host responses and antiviral responses will vary throughout an infection, we set out to quantify *CXCL10*, *CXCL11*, and *TNFSF10* gene expression and variance over the course of SARS-CoV-2 infection by serial NPS sampling over the period of hospitalization. Using an inpatient cohort of 6 subjects (5 females, 1 male – 72.8 years mean age +/-13.3 years), serial samples were collected when study subjects felt well enough to provide a sample (**Figure 2.4A**). *CXCL10*, *CXCL11* and *TNFSF10* transcripts were quantified in all samples provided until the patient was discharged from the inpatient unit (range from 12 to 26 days). Variance for *CXCL10* was 0.213, *CXCL11* was 0.172, and *TNFSF10* was 1.679 (**Figure 2.4B-C**). The variance for each gene was calculated per patient and averaged for each gene, with statistically lower variance observed for *CXCL10* and *CXCL11*, with both of these genes having lower variance than *TNFSF10* ($p < 0.05$). The data from the studies analyzed thus far suggest that *CXCL10* and *CXCL11* are elevated with SARS-CoV-2 infection, track with viral copy number, and are stable throughout the progression of COVID-19 infection. These data suggest that CXCL10 and CXCL11 may be relevant biomarkers that could be useful for tracking respiratory tract viral infections.

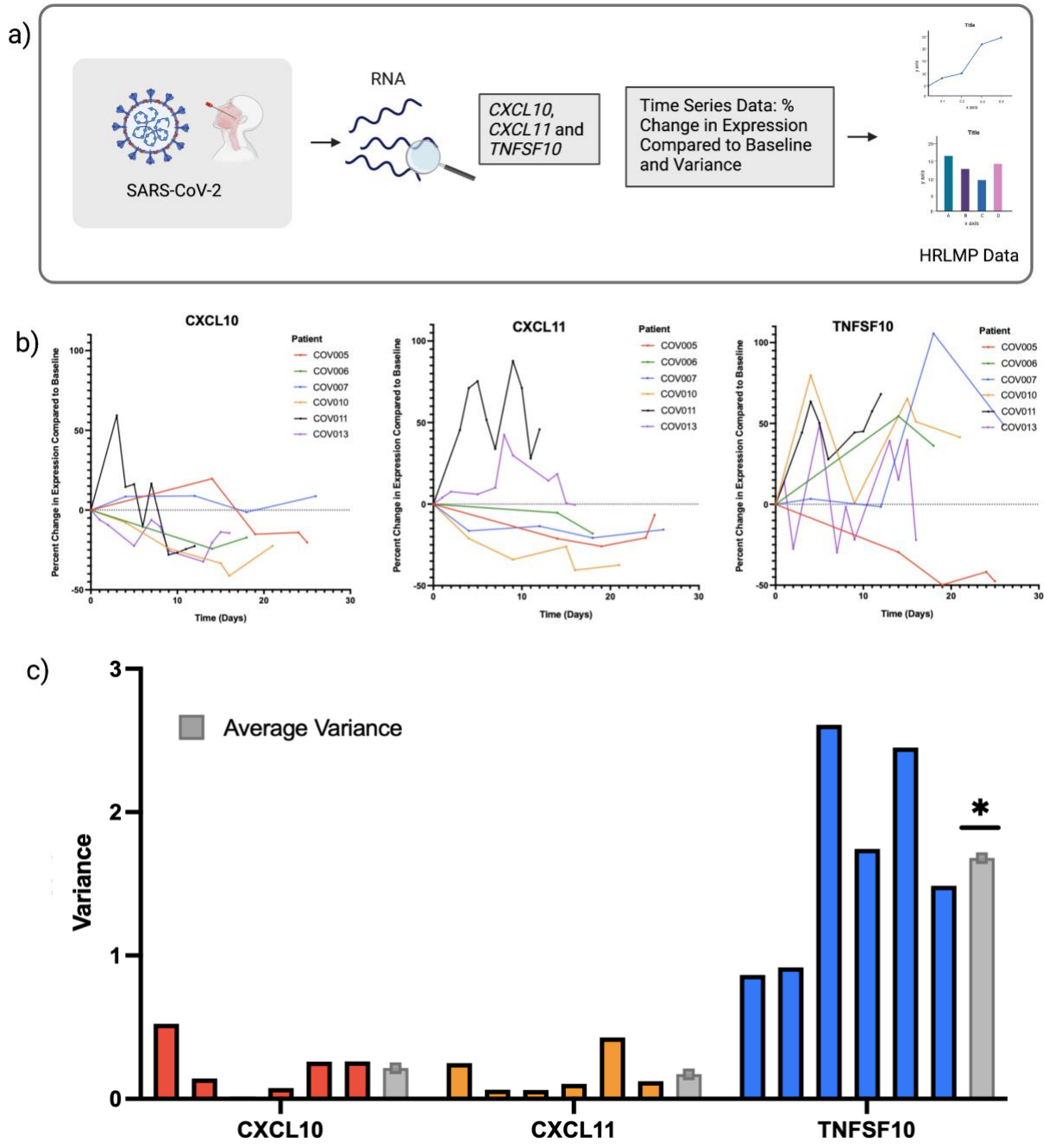
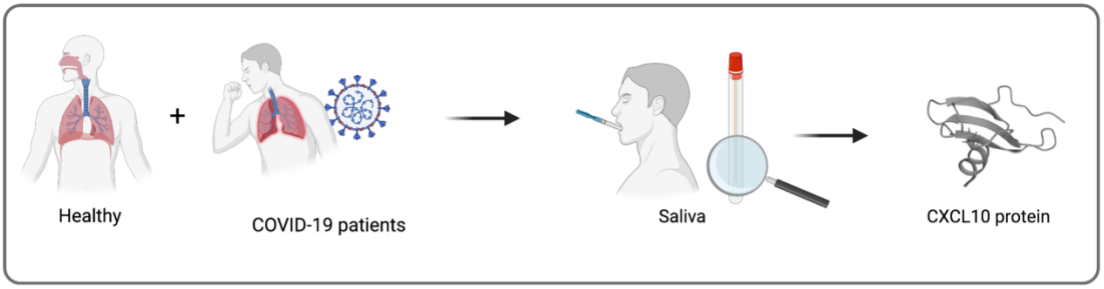


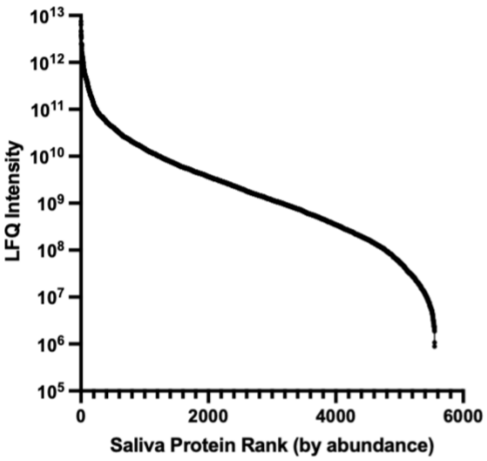

**Figure 2.4:** *CXCL10/CXCL11/TNFSF10* gene signature expression over the course of COVID-19 infection as measured by serial nasopharyngeal swab sampling over period of hospitalization and subsequent daily sampling. **A:** CXCL10. **B:** CXCL11. **C:** TNFSF10. n = 6 independent patients. COV005 = 5 measurements, COV006 = 3 measurements, COV007 = 5 measurements, COV010 = 6 measurements, COV011 = 10 measurements, COV013 = 11 measurements. Data is expressed as % change from first inpatient sampling (set as time=0). To determine which gene fluctuated the least of the course sampling, the variance of RMA values for *CXCL10/CXCL11/TNFSF10* were calculated for each patient and averaged. *CXCL11* mean variance = 0.172, CXCL10 variance = 0.213 and TNFSF10 variance = 1.679).

The data presented identify that *CXCL10* gene demonstrates the greatest magnitude of change during SARS-COV-2 infection in upper airway samples collected via NPS. NPS are not routinely self-administered and present a problematic sampling site for at-home or healthcare resource deficient settings[56]. NPS collected during viral screening are frequently stored in transport medias that are optimized for nucleic acid stabilization and isolation relative to protein stability.[57] Oral sampling has emerged during the COVID-19 pandemic as a useful surrogate for the upper respiratory tract that can be performed by non-healthcare professionals at point of care/need that is amenable to both nucleic acid or protein detection strategies.[58] We therefore explored the utility of detecting CXCL10 protein in oral sampling using saliva collections as a sample format. To be a useful biomarker for monitoring respiratory tract infections, CXCL10 protein should not be detectable in healthy subject samples. Using a mass spectrometry-based proteomics dataset from 8 subjects (4 males and 4 females) that were sampled immediately when waking and again after first meal, we probed the healthy saliva proteome for CXCL10 (**Figure 2.5A**). In a list of 5551 proteins identified by label free quantification (**Figure 2.5B** and **Supplement Data**) CXCL10 protein was not found at any ranking. A table of the 15 most abundant identified proteins lists amylase as the top hit, confirming our analysis pipeline (**Figure 2.5C**). We next quantified the expression of CXCL10 protein in the saliva from a cohort of PCR-confirmed COVID-19 inpatients and healthy un-infected controls. The mean concentration of CXCL10 was 86.4 pg/ml (sd = 109.6) in healthy saliva and 1186.6 pg/ml (sd= 1252.3) in COVID-19 inpatients (**Figure 2.5D**, $p < 0.05$).

a)




b)




c)

| Rank | Protein names | Gene names |
|---|---|---|
| 1 | Alpha-amylase 1 | AMY1A |
| 2 | Annexin A1;Annexin | ANXA1 |
| 3 | Desmoplakin | DSP |
| 4 | Serum albumin | ALB |
| 5 | Protein S100-A9 | S100A9 |
| 6 | Actin, cytoplasmic 2;Actin, cytoplasmic 2, N-terminally processed | ACTG1 |
| 7 | Neuroblast differentiation-associated protein AHNAK | AHNAK |
| 8 | Ig alpha-1 chain C region | IGHA1 |
| 9 | Zymogen granule protein 16 homolog B | ZG16B |
| 10 | Alpha-2-macroglobulin-like protein 1 | A2ML1 |
| 11 | Small proline-rich protein 3 | SPRR3 |
| 12 | Polymeric immunoglobulin receptor;Secretory component | PIGR |
| 13 | Periplakin | PPL |
| 14 | Protein S100-A8;Protein S100-A8, N-terminally processed | S100A8 |
| 15 | Cornulin | CRNN |

d)

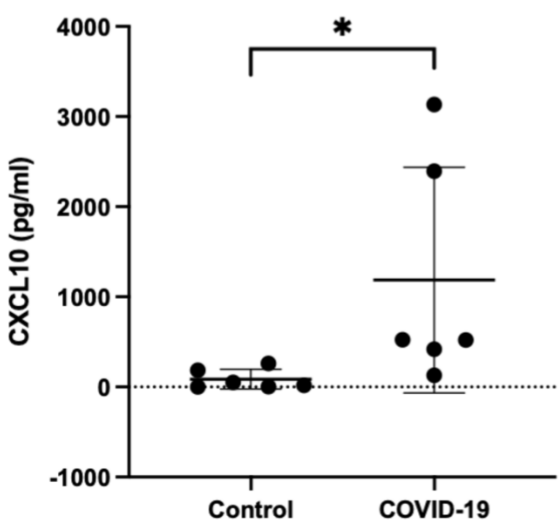



**Figure 2.5:** CXCL10 protein is not present in healthy saliva samples and elevated during infection in saliva from COVID-19 patients**. A:** From Grassl et al[49] ultra deep analysis of the healthy saliva proteome, the CXCL10 protein was not found amongst the list of 5562 proteins (Supplementary data). **B:** CXCL10 Levels in Saliva of SARS-CoV-2 in Hospitalized COVID-19 patients quantified in using Human Cytokine Array / Chemokine Array 71-plex (Eve Technologies, Calgary, Alberta, Canada). Healthy volunteers were used as the control group. Mean concentration of CXCL10 was 86.43 pg/ml (sd = 109.6, n=6) in healthy saliva, and mean of 1186.65 pg/ml sd= 1252.34 in Covid-19 patients. COVID-19 group showed a significantly greater CXCL10 concentration ($p < 0.05$). CXCL10 Protein Structures generated using https://www.uniprot.org/.

The low levels of CXCL10 protein in saliva of healthy subjects but elevated levels in COVID-19 subjects is amenable to conventional LFA development that detects the upregulation of a mediator of interest. LFAs have become widely accepted tools for self-testing during COVID-19 and represent a technology platform useful for at-home and healthcare deficient settings. Therefore, we next pursued the development of an open-source CXCL10 LFA using commercially available reagents (**Figure 2.6A**). A recombinant anti-human CXCL10 mouse IgG-monoclonal and anti-mouse IgG goat-polyclonal were selected as an antibody pair for CXCL10 detection. Assay development was validated in ideal buffer demonstrating a sensitivity of 2ng/ml (**Figure 2.6B-C**). Assay testing in commercially available artificial saliva reproduced the sensitivity of 2ng/ml (**Figure 2.6D-E**). Lastly, the assay detected 10ng/ml CXCL10 protein spiked into real human saliva from healthy controls that was not detected in control /non-spiked samples from the same donors (**Figure 2.6F**). These results demonstrate the sensitivity of an open-source CXCL10 LFA prototype for self-administered point of care saliva testing with potential applications in respiratory tract viral infection monitoring.

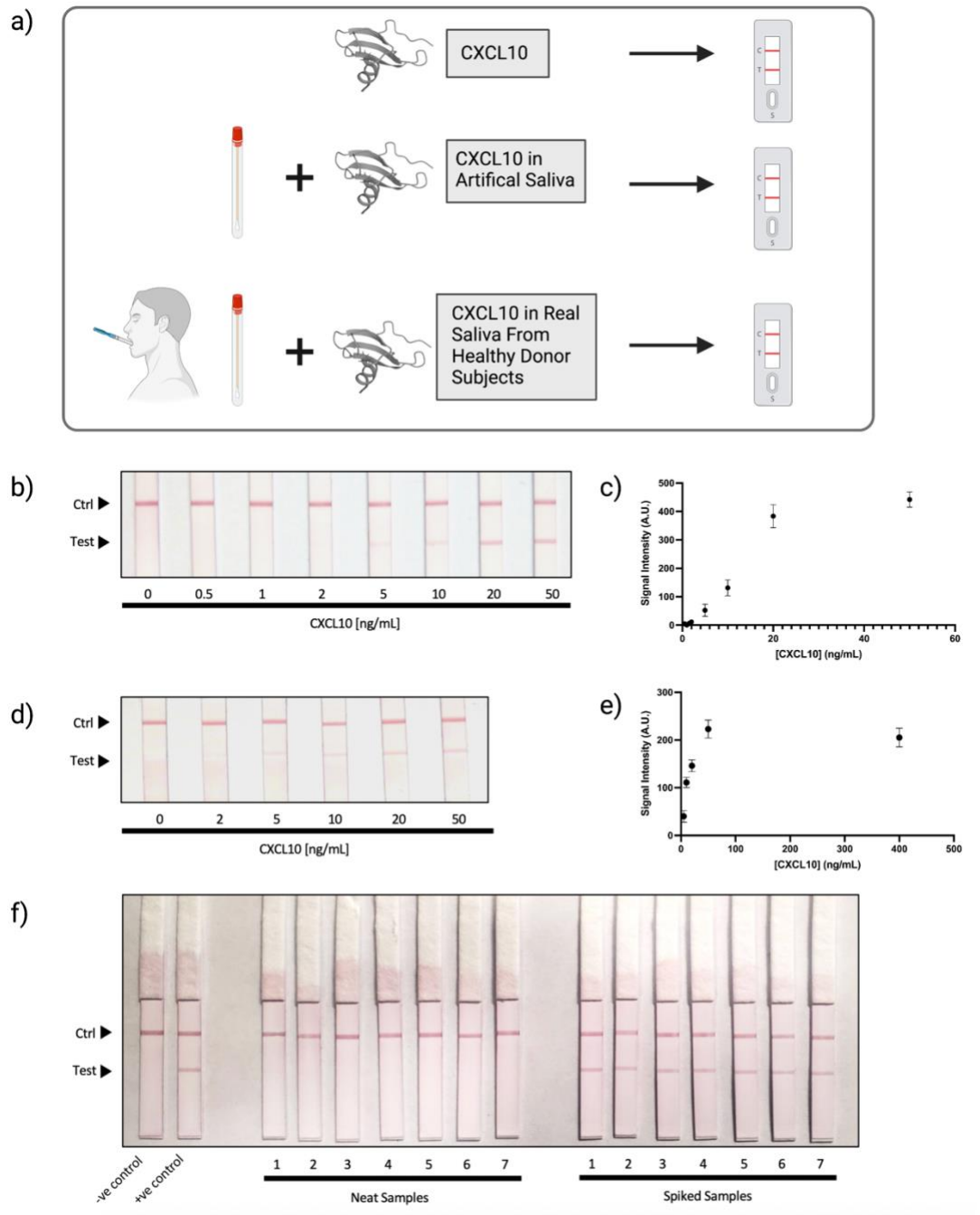



**Figure 2.6:** CXCL10 protein detection is feasible in fresh human saliva

**A**: Overview of development and testing

**B-C**: Sensitivity development and validation in ideal buffer (see methods). Limit of detection of CXCL10 using a hand-help reader. A positive signal is generated at 2ng/mL with n=5.

**D-E**: Sensitivity test in artificial saliva (see methods)

**F**: Real world testing in human saliva from healthy control without (neat) and with (spiked) CXCL10 addition.

CXCL10 Protein Structures generated using https://www.uniprot.org/.

## 2.4 Discussion

ISGs play a key role in defence against viral infections and measuring their expression levels during respiratory infections may provide diagnostic and prognostic information beyond measurement of a pathogen alone. In the present paper, we characterize three ISGs of interest that result in secreted protein products amenable to extracellular detection strategies: *CXCL10*, *CXCL11* and *TNFSF10.* We quantify *CXCL10, CXCL11* and *TNFSF10* expression levels in the context of multiple respiratory tract viral pathogens including RSV, RV, influenza A and SARS-CoV-2. Gene expression levels from upper airway samples suggested that *CXCL10* elevations were consistent across multiple viruses, had qualitative positive trends with measures of severity of infection (e.g., viral load), and had the lowest variance over course of COVID-19 infection of the three candidates examined. To transition from upper airway samples collected by a healthcare professional to a self-collected sample format, we next explored CXCL10 levels in oral samples. Deep proteomic profiling of healthy human saliva suggested low to absent levels of CXCL10 protein in healthy subjects. In contrast, saliva from SARS-CoV-2 infected individuals resulted in elevations in CXCL10. Leveraging the relative binary behaviour of CXCL10 in healthy vs SARS-CoV-2 infected sample groups, we developed a prototype LFA for CXCL10 protein with a sensitivity of 2ng/ml in human saliva. Collectively, our work adds to the growing importance of examining host biomarkers during respiratory tract infections for diagnostic and prognostic value and demonstrates the feasibility of using self-administered qualitative LFA detection methods.

Our study used a combination of publicly available datasets and *de novo* generated datasets. Consequently, the results are heavily dependent on the annotation of the data provided for each study. Differences in study design and the definition of control groups may contribute to potential biases and limitations in the data. Commercially available or bespoke PCR-based respiratory panels were used in several of the studies to annotate positive and negative subjects in individuals suspected of respiratory tract viral infection. PCR-based detection strategies can diagnose several viruses simultaneously while having greater challenges for diagnosing bacterial respiratory infections.[59] The challenges with bacterial diagnostic decisions in the respiratory tract are rooted into the commensal nature of bacteria in the nasopharynx such as *Streptococcus pneumoniae*[60] and *Haemophilus influenzae*.[61] For example, up to 10% of adults are carriers of *Streptococcus pneumoniae*,[60] creating a challenge to determine whether this organism is causing pathology or a part of the individual's microbiome. The potential for viral and bacterial co-infections or presence of complex commensal communities can create limitations in analyzing data where only viral infection status is provided. Any underlying virus or bacterial respiratory tract infection would impact host gene expression patterns that may include *CXCL10, CXCL11,* and *TNFSF10*. Furthermore, studies focused only on a single virus may not report the outcomes for other viruses detected in a multi-plex PCR assay. The publicly available COVID-19 negative labelled samples were generated during routine public SARS-CoV-2 testing campaigns. As a result, although an individual may be labelled as SARS-CoV-2 negative, they may have had an underlying viral or bacterial infection not quantified and recorded. Although the incidence of influenza A was low during COVID-19,[62,63] there remained the possibility that other respiratory tract viral or bacterial infections were present in individuals that tested negative for SARS-CoV-2.[40] In Lieberman et al[35], patients suspected of SARS-CoV-2 infection were confirmed by RT-

PCR result while those testing negative became the controls although no confirmations for other viral or bacterial respiratory tract infections were performed. Mick et al[48] took a distinct approach by using metagenomic RNA-sequencing and separating their cohort into three groups – 1) *SARS-CoV-2 positive* - no other viral RTI 2) *SARS-CoV-2 negative* - positive for another virus and 3) *SARS-CoV-2 negative* - no other RTI viruses detected. Metagenomic RNA-sequencing can simultaneously identify host genes and microbial RNA and can allow for an unbiased approach to labelling control groups and create further transparency when sharing data publicly.[64] Metagenomic molecular diagnostic strategies have been optimized for upper airway samples and are likely to be useful strategies moving forward with host-pathogen diagnostic approaches and be informative for future pandemics.[65]

A variety of publicly available datasets and databases exist and are increasing in accessibility. Notable examples of databases include the NCBI resources Gene Expression Omnibus, the Broad Institute's Connectivity Map, and the Cancer Genome Atlas. The diversity of these databases has been consolidated in meta-databases to attempt to harmonize publicly available resources. Harmonizome is a publicly accessible meta-database of 112 datasets from 65 resources (https://maayanlab.cloud/Harmonizome/about). We leveraged the consolidation performed by Harmonizome to determine the expression of *CXCL10*, *CXCL11*, and *TNFSF10* across transcriptomic datasets defined by "GEO Signatures of Differentially Expressed Genes for Viral Infections". Annotation of the datasets under this search term revealed some non-human datasets and datasets derived from non-respiratory tract viral infections, which were subsequently removed to maintain focus on host-biomarkers to respiratory tract infections. A limitation of the datasets curated by Harmonizome is that they are pre-clinical studies that include multiple different viruses from multiple different cell-lines in addition to different time-points post-infection. Despite the potential for extensive variation in experimental conditions between datasets, a significant signal for an increase in *CXCL10*, *CXCL11*, and *TNFSF10* was observed across the meta-database, without observing the same for three common housekeeping genes, *ACTB*, *GADPH* and *TUBB*. Although Harmonizome contains multiple datasets and databases, an absence of well-annotated samples from respiratory tract viral infections in humans required additional dataset generation and curation to extend and validate observations of *CXCL10*, *CXCL11*, and *TNFSF10* elevations.

NPS and mid-turbinate swabs are clinical standard sampling strategies for analysis of respiratory tract viral infections and are amenable to PCR and transcriptomic based readouts.[56,65,66] Historically, the small sample amounts collected from upper airway swabs have limited the ability to perform transcriptomic analysis which resulted in a focus on multi-plex PCR technologies.[67] The development of microarray technologies for small amounts of input material and the reduction in sequencing costs that affords greater sequencing depth have increased accessibility to transcriptomics. Prior to COVID-19 pandemic, relatively few studies of respiratory tract viral infections capitalized on technological advances for transcriptomics on upper airway samples. A seminal study demonstrating the ability to perform unbiased transcriptomics on upper airway samples was performed on a pediatric cohort of individuals with suspected and subsequent confirmation of respiratory tract infection via PCR.[47] We explored this publicly available host transcriptomic dataset from a cohort of 24 subjects. Independent of data that defines the time of initial infection, *CXCL10* and *CXCL11* up-regulation was positively correlated with RSV infection when compared to control. In contrast to RSV, no significant

differences were observed for *CXCL10, CXCL11* and *TNFSF10* when comparing upper airway swabs from RV infected individuals to controls. The small sample size of this study precludes drawing conclusions about the absence of signal for our candidate host biomarkers with RV infection. Related to sample sizes of publicly available host-transcriptomic datasets from upper airway samples, the COVID-19 pandemic catalyzed the mass adoption of molecular sequencing technologies and has subsequently demonstrated the feasibility and utility of taking this approach, laying the foundations for a future state of deep host-pathogen diagnostics.[35,40,48,65] A limitation of leveraging clinical samples taken during diagnostic processes for research purposes is that there is lack of control over the time of initial infection, which could impact host gene expression signatures, particularly of ISGs which are time-sensitive.[23] This limitation may be overcome in the future with specific research study designs that could include prospectively following cohorts in a surveillance design to capture natural infections [67] or controlled human studies with viral challenges [68] with informed sample size calculations made based on the emerging publicly available datasets.

The presence of an elevated host biomarker during a confirmed respiratory tract viral infection may provide utility in determining stratifying those individuals at risk of morbidity and mortality, which could help inform treatment options at the time of first diagnosis.[25,26] To explore the possibility that either *CXCL10, CXCL11* and *TNFSF10* could function as a prognostic biomarker for disease severity, we explored expression levels in the context of SARS-CoV-2 viral copy number and COVID-19 mortality as datasets from other respiratory tract viral infections and viral copy are not publicly available. *CXCL10* and *CXCL11* have been shown to significantly increase in expression in NPS samples from SARS-CoV-2 patients compared with control[69]. Using a large publicly available host transcriptomic dataset with PCR confirmed SARS-CoV-2 positive cases and reported Ct values, *CXCL10, CXCL11* and *TNFSF10* are quantitatively observed to correlate with viral copy number (https://covidgenes.weill.cornell.edu/). The correlation between SARS-CoV-2 viral load in nasal samples tracking with severe disease has been supported in most but not all cases. In a prospective study, lower Ct values (higher viral load) correlated with severe disease in hospitalized patients. Although, there was no significant correlation between lower Ct values (higher viral load) for the risk of being hospitalized for SARS-COV-2 when adjusted for the time of symptoms onset.[54] A longitudinal study found increased SARS-CoV-2 viral load correlated with a higher risk of death, and increased inflammatory markers CRP and IL-6.[55] Conversion of viral copy Ct value to a continuous variable (>40 = 0, <15 = 1) resulted in a statistically significant correlation between *CXCL10, CXCL11* and *TNFSF10* and amount of virus present. Despite the reports that SARS-CoV-2 viral copy number may relate to severity of COVID-19 disease, differences in the measure of severity and the populations chosen in independent studies have contributed to ambiguous conclusions. In addition to viral copy number, host responses have been measured throughout the COVID-19 pandemic and have been demonstrated to have predictive value for determining severity. Indeed, host responses to SARS-CoV-2 infection including serum IL-6 and C-Reactive Protein have been evaluated and demonstrated prognostic value in triaging COVID-19 patients.[70,71] In parallel to analyses relating viral copy number to host responses, we investigated death as an unambiguous measure of severity. In a cohort of 25 intensive care unit admitted subjects, each with PCR-confirmed SARS-CoV-2 infection, the upper airway swab sample taken at time of diagnosis showed no significant differences in magnitude of *CXCL10, CXCL11* and *TNFSF10* gene transcript between fatal and non-fatal

COVID-19 cases. Previous research in blood samples of COVID-19 patients looked at 53 potential biomarkers and found CXCL10 as the best predictor of death.[37] They also observed that CXCL10 levels were increased in patients with ICU care compared to without ICU and CXCL10 levels decreased in COVID-19 survivors who were discharged from the hospital.[37] This study used a different sample type, blood, in addition, to determine CXCL10 levels, they used Bio-Plex Pro™ Human Cytokine 27-plex, which determined CXCL10 protein concentrations. In addition to Lorè et al[37], several other studies using blood samples found CXCL10 was associated with COVID-19 disease severity.[72,73] These data suggest that focusing on blood as the sample type may provide a greater differentiating signature than upper airway gene transcripts. To date, point of care blood sampling for host biomarkers has received little traction in the infectious disease space. In contrast, self-administered LFAs for detection of protein in oral and upper airway samples provide a strategy for implementing point of care monitoring of host biomarkers relevant in respiratory tract infections.

The significant limitation of being blind to the time-point of initial infection is intrinsic to cross-sectional studies performed on individuals during initial diagnosis. To overcome this limitation, time-series studies may be performed to capture natural infections in the community, or the time course of infection can be monitored once an individual is admitted to the healthcare system, with admission becoming a baseline for the given study subject.[67] We took this secondary approach of defining baseline as the time of hospital admission and analyzed the variance in *CXCL10, CXCL11* and *TNFSF10* over the course of hospitalization. Analysis of variance for each biomarker in study subjects repeatedly sampled during their hospitalization was completed with each individual's expression levels at admission functioning as their baseline. We quantified variance for each gene over time to identify which candidate may function as a more robust measure of host response to SARS-CoV-2 infection that is independent of the time of first sampling and diagnosis. The observation that *CXCL10* and *CXCL11* had lower variance over *TNFSF10* supported excluding the latter as a robust biomarker stable throughout the course of an infection. The stability of *CXCL10* and *CXCL11* that we observed in upper airway samples has been observed in serial blood and serum samples from COVID-19 patients.[72,73] Previous *in-vitro* studies found CXCL10 and CXCL11 expression in SARS-CoV-2 infected Calu-3 human lung cells peaked at 72 hours post infection.[39] CXCL10 showed greater log fold change and remained significantly elevated from 12-72 hours post-infection and CXCL11 elevated 18-72 hours post-infection.[39] Collectively, these data support the utility of CXCL10 protein as a prognostic biomarker for COVID-19 and provided a rationale for our focus on this mediator for downstream LFA development for self-administered testing.

Self-administered saliva testing for respiratory tract viral infections has become a reality through the COVID-19 pandemic.[58] The qualitative nature of LFA has been challenged in the popular and academic circles [74(p20)] with the net result being the widespread adoption of the technology on a global scale to provide a degree of diagnostic capacity previously not available to the general public. The possibility that other saliva-based tests relevant in infectious disease may emerge has therefore increased and may include host biomarkers of clinical importance.[75] The qualitative nature of LFAs requires a clear binary response of presence and absence for the test to be practical. Our *a priori* decision to characterize *CXCL10, CXCL11*, and *TNFSF10* were based on the assumption of their extracellular secretion with infection and low levels during times of no infection. Using a deep proteomic profiling of healthy human saliva from males and females

with repeated measures, CXCL10 was not identified in a list of 5551 proteins detected with 95% of the proteins identified in at least 3 samples and nearly 80% across all 8 samples. In contrast to deep proteomic profiling, we used a multiplex cytokine array from Eve Technologies and were able to detect CXCL10 with a mean concentration of 86.43 pg/ml (sd = 109.6, n = 6) in saliva from healthy subjects. Although these data may appear in conflict, an alternative explanation is that even deep proteomic profiling may not achieve the sensitivity of commercially available established multiplex cytokine arrays. Importantly, an elevation in CXCL10 protein was observed when analyzing saliva samples from COVID-19 infected subjects. The observed elevation of saliva CXCL10 protein provides a rationale to explore LFA technology for point of care detection and provides guidance on the sensitivity requirements for distinguishing samples from negative non-infected individuals and positive SARS-CoV-2 infected individuals. Leveraging commercially available reagents for LFA development including detection and capture antibodies, nitrocellulose membranes, and absorbent wicking material, we developed a prototype CXCL10 protein detection assay with a sensitivity of 2 ng/ml in artificial and healthy saliva. To our knowledge, this was the first CXCL10 LFA designed for human saliva samples. A previous study developed an CXCL10 LFA for amniotic fluid as a marker of intra-amniotic inflammation with sensitivity of 100 pg/ml (0.1 ng/ml).[76] This study was able to achieve greater sensitivity, suggesting that the prototype presented could be optimized for detection of CXCL10 levels relevant during respiratory tract infections but not during periods of health. Importantly, if the sensitivity of a CXCL10 LFA were to be reduced to a level that can detect protein in healthy subjects, the utility of the test would be reduced as it would not help differentiate elevated levels of CXCL10 that may predict COVID-19 severity. Consistent with this requirement, CXCL10 protein levels peaked at around 60 ng/ml on day 2 of RV infection in adult nasal lavages samples which is within the range of our prototype CXCL10 LFA.[33] It was also observed the concentration of CXCL10 measured with ELISA on the day of intubation from infants with RSV to be around 33 ng/ml from bronchoalveolar lavage samples [31] which is also within the range of our CXCL10 LFA.

Our characterization of CXCL10 gene expression in several datasets and development of a prototype CXCL10 responsive LFA amenable to saliva sampling represent first steps in a process towards biomarker validation and demonstration of clinical utility. Future utility could include examining the ability of CXCL10 protein levels in saliva to aid in discrimination of viral vs. bacterial respiratory tract infections, which would aid in rapid decision making for antibiotic administration as part of antibiotic stewardship practices. Future studies from clinically phenotyped subjects with respiratory tract bacterial infections in the absence of viral infections or bacterial/viral co-infections are required for this important research question to be answered. Perhaps closer to the clinic due could be the for the prognosis of COVID-19 during SARS-CoV-2 infection. Previous research has found CXCL10 to play key role in predicting ICU admission using blood samples.[37,72,73] To validate CXCL10 in saliva as a biomarker of SARS-CoV-2 severity, further longitudinal research into time series data would be required. This would need to prove that CXCL10 levels are increased in a time frame where appropriate intervention is possible. It would not be useful to measure CXCL10 to predict severity if there are other more obvious clinical symptoms already present.

To conclude, we present data that characterizes *CXCL10*, *CXCL11* and *TNFSF10* gene expression in upper airway samples from individuals with respiratory tract viral infections. We provide a justification that CXCL10 protein product is amenable to LFA testing in a point of care setting. Further clinical validation of CXCL10 gene and protein as a biomarker of respiratory tract viral infections is required and should be complemented with study cohorts that include bacterial infections in the absence of viral infections. The utility of measuring CXCL10 at the point of care for diagnostic and prognostic purposes will grow as the appreciation that both host and pathogen molecular markers are important for optimal clinical care and healthcare system utilization.

**2.5 References:**

1. World Health Organization. Global Health Estimates 2020: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2019. Published 2020. Accessed July 28, 2022. https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death

2. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*. 2020;20(5):533-534. doi:10.1016/S1473-3099(20)30120-1

3. Hu B, Guo H, Zhou P, Shi ZL. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol*. 2021;19(3):141-154. doi:10.1038/s41579-020-00459-7

4. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*. 2012;380(9859):2095-2128. doi:10.1016/S0140-6736(12)61728-0

5. Fendrick AM, Monto AS, Nightengale B, Sarnes M. The economic burden of non-influenza-related viral respiratory tract infection in the United States. *Arch Intern Med*. 2003;163(4):487-494. doi:10.1001/archinte.163.4.487

6. European Respiratory Society. *European Lung White Book: The Economic Burden of Lung Disease*. European Respiratory Society; 2011. Accessed July 26, 2022. https://www.erswhitebook.org/chapters/the-economic-burden-of-lung-disease/the-cost-of-respiratory-disease/

7. World Economic Outlook, October 2021: Recovery During A Pandemic. IMF. Accessed April 25, 2022. https://www.imf.org/en/Publications/WEO/Issues/2021/10/12/world-economic-outlook-october-2021

8. Cutler DM, Summers LH. The COVID-19 Pandemic and the $16 Trillion Virus. *JAMA*. 2020;324(15):1495-1496. doi:10.1001/jama.2020.19759

9. Chowdhury P, Paul SK, Kaisar S, Moktadir MdA. COVID-19 pandemic related supply chain studies: A systematic review. *Transp Res Part E Logist Transp Rev*. 2021;148:102271. doi:10.1016/j.tre.2021.102271

10. Krammer F, Smith GJD, Fouchier RAM, et al. Influenza. *Nat Rev Dis Primer*. 2018;4(1):3. doi:10.1038/s41572-018-0002-y

11. Pavia AT. Viral infections of the lower respiratory tract: old viruses, new viruses, and the role of diagnosis. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2011;52 Suppl 4:S284-289. doi:10.1093/cid/cir043

12. Lindemans CA, Kimpen JLL. The Immune Response to Viral Lower Respiratory Tract Infection. *Hot Top Infect Immun Child II*. 2005;568:41-65. doi:10.1007/0-387-25342-4_4

13. Paget J, Spreeuwenberg P, Charu V, et al. Global mortality associated with seasonal influenza epidemics: New burden estimates and predictors from the GLaMOR Project. *J Glob Health*. 9(2):020421. doi:10.7189/jogh.09.020421

14. Jacobs SE, Lamson DM, St. George K, Walsh TJ. Human Rhinoviruses. *Clin Microbiol Rev*. 2013;26(1):135-162. doi:10.1128/CMR.00077-12

15. Papadopoulos NG, Papi A, Psarras S, Johnston SL. Mechanisms of rhinovirus-induced asthma. *Paediatr Respir Rev*. 2004;5(3):255-260. doi:10.1016/j.prrv.2004.04.002

16. Borchers AT, Chang C, Gershwin ME, Gershwin LJ. Respiratory Syncytial Virus—A Comprehensive Review. *Clin Rev Allergy Immunol*. 2013;45(3):331-379. doi:10.1007/s12016-013-8368-9

17. Jozwik A, Habibi MS, Paras A, et al. RSV-specific airway resident memory CD8+ T cells and differential disease severity after experimental human infection. *Nat Commun*. 2015;6:10224. doi:10.1038/ncomms10224

18. Glezen WP, Taber LH, Frank AL, Kasel JA. Risk of primary infection and reinfection with respiratory syncytial virus. *Am J Dis Child 1960*. 1986;140(6):543-546. doi:10.1001/archpedi.1986.02140200053026

19. Shi T, McAllister DA, O'Brien KL, et al. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. *Lancet Lond Engl*. 2017;390(10098):946-958. doi:10.1016/S0140-6736(17)30938-8

20. Makrinioti H, Hasegawa K, Lakoumentas J, et al. The role of respiratory syncytial virus- and rhinovirus-induced bronchiolitis in recurrent wheeze and asthma-A systematic review and meta-analysis. *Pediatr Allergy Immunol Off Publ Eur Soc Pediatr Allergy Immunol*. 2022;33(3):e13741. doi:10.1111/pai.13741

21. Knudson CJ, Varga SM. The Relationship Between Respiratory Syncytial Virus and Asthma. *Vet Pathol*. 2015;52(1):97-106. doi:10.1177/0300985814520639

22. Moriyama M, Hugentobler WJ, Iwasaki A. Seasonality of Respiratory Viral Infections. *Annu Rev Virol*. 2020;7(1):83-101. doi:10.1146/annurev-virology-012420-022445

23. Mostafavi S, Yoshida H, Moodley D, et al. Parsing the Interferon Transcriptional Network and Its Disease Associations. *Cell*. 2016;164(3):564-578. doi:10.1016/j.cell.2015.12.032

24. Schoggins JW. Interferon-Stimulated Genes: What Do They All Do? *Annu Rev Virol*. 2019;6(1):567-584. doi:10.1146/annurev-virology-092818-015756

25. Ashkenazi-Hoffnung L, Oved K, Navon R, et al. A host-protein signature is superior to other biomarkers for differentiating between bacterial and viral disease in patients with respiratory infection and fever without source: a prospective observational study. *Eur J Clin Microbiol*

*Infect Dis Off Publ Eur Soc Clin Microbiol*. 2018;37(7):1361-1371. doi:10.1007/s10096-018-3261-3

26. Gupta RK, Rosenheim J, Bell LC, et al. Blood transcriptional biomarkers of acute viral infection for detection of pre-symptomatic SARS-CoV-2 infection: a nested, case-control diagnostic accuracy study. *Lancet Microbe*. 2021;2(10):e508-e517. doi:10.1016/S2666-5247(21)00146-4

27. Newton AH, Cardani A, Braciale TJ. The host immune response in respiratory virus infection: balancing virus clearance and immunopathology. *Semin Immunopathol*. 2016;38(4):471-482. doi:10.1007/s00281-016-0558-0

28. Mesev EV, LeDesma RA, Ploss A. Decoding type I and III interferon signalling during viral infection. *Nat Microbiol*. 2019;4(6):914-924. doi:10.1038/s41564-019-0421-x

29. Guo-Parke H, Canning P, Douglas I, et al. Relative Respiratory Syncytial Virus Cytopathogenesis in Upper and Lower Respiratory Tract Epithelium. *Am J Respir Crit Care Med*. 2013;188(7):842-851. doi:10.1164/rccm.201304-0750OC

30. Lu X, Masic A, Liu Q, Zhou Y. Regulation of influenza A virus induced CXCL-10 gene expression requires PI3K/Akt pathway and IRF3 transcription factor. *Mol Immunol*. 2011;48(12):1417-1423. doi:10.1016/j.molimm.2011.03.017

31. McNamara PS, Flanagan BF, Hart CA, Smyth RL. Production of Chemokines in the Lungs of Infants with Severe Respiratory Syncytial Virus Bronchiolitis. *J Infect Dis*. 2005;191(8):1225-1232. doi:10.1086/428855

32. Liu M, Guo S, Hibbert JM, et al. CXCL10/IP-10 in infectious diseases pathogenesis and potential therapeutic implications. *Cytokine Growth Factor Rev*. 2011;22(3):121-130. doi:10.1016/j.cytogfr.2011.06.001

33. Spurrell JCL, Wiehler S, Zaheer RS, Sanders SP, Proud D. Human airway epithelial cells produce IP-10 (CXCL10) in vitro and in vivo upon rhinovirus infection. *Am J Physiol-Lung Cell Mol Physiol*. 2005;289(1):L85-L95. doi:10.1152/ajplung.00397.2004

34. Machado D, Hoffmann J, Moroso M, et al. RSV Infection in Human Macrophages Promotes CXCL10/IP-10 Expression during Bacterial Co-Infection. *Int J Mol Sci*. 2017;18(12):2654. doi:10.3390/ijms18122654

35. Lieberman NAP, Peddu V, Xie H, et al. In vivo antiviral host transcriptional response to SARS-CoV-2 by viral load, sex, and age. Cadwell K, ed. *PLOS Biol*. 2020;18(9):e3000849. doi:10.1371/journal.pbio.3000849

36. Peiris J, Yu W, Leung C, et al. Re-emergence of fatal human influenza A subtype H5N1 disease. *Lancet Lond Engl*. 2004;363(9409):617-619. doi:10.1016/S0140-6736(04)15595-5

37. Lorè NI, De Lorenzo R, Rancoita PMV, et al. CXCL10 levels at hospital admission predict COVID-19 outcome: hierarchical assessment of 53 putative inflammatory biomarkers in an observational study. *Mol Med*. 2021;27(1):129. doi:10.1186/s10020-021-00390-4

38. Metzemaekers M, Vanheule V, Janssens R, Struyf S, Proost P. Overview of the Mechanisms that May Contribute to the Non-Redundant Activities of Interferon-Inducible CXC Chemokine Receptor 3 Ligands. *Front Immunol*. 2018;8. Accessed April 10, 2022. https://www.frontiersin.org/article/10.3389/fimmu.2017.01970

39. Callahan V, Hawks S, Crawford MA, et al. The Pro-Inflammatory Chemokines CXCL9, CXCL10 and CXCL11 Are Upregulated Following SARS-CoV-2 Infection in an AKT-Dependent Manner. *Viruses*. 2021;13(6):1062. doi:10.3390/v13061062

40. Butler D, Mozsary C, Meydan C, et al. Shotgun transcriptome, spatial omics, and isothermal profiling of SARS-CoV-2 infection reveals unique host responses, viral diversification, and drug interactions. *Nat Commun*. 2021;12(1):1660. doi:10.1038/s41467-021-21361-7

41. Abdulghani J, El-Deiry WS. TRAIL receptor signaling and therapeutics. *Expert Opin Ther Targets*. 2010;14(10):1091-1108. doi:10.1517/14728222.2010.519701

42. Cummins N, Badley A. The TRAIL to viral pathogenesis: the good, the bad and the ugly. *Curr Mol Med*. 2009;9(4):495-505. doi:10.2174/156652409788167078

43. Koczula KM, Gallotta A. Lateral flow assays. *Essays Biochem*. 2016;60(1):111-120. doi:10.1042/EBC20150012

44. Lau CS, Hoo SP, Koh JMJ, Phua SK, Aw TC. Performance of the Roche IL-6 chemiluminescent immunoassay in patients with COVID-like respiratory symptoms. *J Virol Methods*. 2021;296:114224. doi:10.1016/j.jviromet.2021.114224

45. FDA. Elecsys IL-6 Package Insert 2020-06, V1.0; Material Numbers 05109442190 and 07027532190.

46. Rouillard AD, Gundersen GW, Fernandez NF, et al. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database*. 2016;2016(baw100). doi:10.1093/database/baw100

47. Yu J, Peterson DR, Baran AM, et al. Host Gene Expression in Nose and Blood for the Diagnosis of Viral Respiratory Infection. *J Infect Dis*. 2019;219(7):1151-1161. doi:10.1093/infdis/jiy608

48. Mick E, Kamm J, Pisco AO, et al. Upper airway gene expression reveals suppressed immune responses to SARS-CoV-2 compared with other respiratory viruses. *Nat Commun*. 2020;11(1):5854. doi:10.1038/s41467-020-19587-y

49. Grassl N, Kulak NA, Pichler G, et al. Ultra-deep and quantitative saliva proteome reveals dynamics of the oral microbiome. *Genome Med*. 2016;8(1):1-13. doi:10.1186/s13073-016-0293-0

50. Schaab C, Geiger T, Stoehr G, Cox J, Mann M. Analysis of high accuracy, quantitative proteomics data in the MaxQB database. *Mol Cell Proteomics MCP*. 2012;11(3):M111.014068. doi:10.1074/mcp.M111.014068

51. Tyanova S, Temu T, Sinitcyn P, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*. 2016;13(9):731-740. doi:10.1038/nmeth.3901

52. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30(1):207-210. doi:10.1093/nar/30.1.207

53. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res*. 2013;41(Database issue):D991-995. doi:10.1093/nar/gks1193

54. Knudtzen FC, Jensen TG, Lindvig SO, et al. SARS-CoV-2 viral load as a predictor for disease severity in outpatients and hospitalised patients with COVID-19: A prospective cohort study. *PLOS ONE*. 2021;16(10):e0258421. doi:10.1371/journal.pone.0258421

55. Fajnzylber J, Regan J, Coxen K, et al. SARS-CoV-2 viral load is associated with increased disease severity and mortality. *Nat Commun*. 2020;11(1):5493. doi:10.1038/s41467-020-19057-5

56. Smieja M, Castriciano S, Carruthers S, et al. Development and evaluation of a flocked nasal midturbinate swab for self-collection in respiratory virus infection diagnostic testing. *J Clin Microbiol*. 2010;48(9):3340-3342. doi:10.1128/JCM.02235-09

57. Luinstra K, Petrich A, Castriciano S, et al. Evaluation and clinical validation of an alcohol-based transport medium for preservation and inactivation of respiratory viruses. *J Clin Microbiol*. 2011;49(6):2138-2142. doi:10.1128/JCM.00327-11

58. Wyllie AL, Fournier J, Casanovas-Massana A, et al. Saliva or Nasopharyngeal Swab Specimens for Detection of SARS-CoV-2. *N Engl J Med*. 2020;383(13):1283-1286. doi:10.1056/NEJMc2016359

59. Ramilo O, Mejias A. Host transcriptomics for diagnosis of infectious diseases: one step closer to clinical application. *Eur Respir J*. 2017;49(6). doi:10.1183/13993003.00993-2017

60. Weiser JN, Ferreira DM, Paton JC. Streptococcus pneumoniae: transmission, colonization and invasion. *Nat Rev Microbiol*. 2018;16(6):355-367. doi:10.1038/s41579-018-0001-8

61. Zhu H, Wang A, Tong J, et al. Nasopharyngeal carriage and antimicrobial susceptibility of Haemophilus influenzae among children younger than 5 years of age in Beijing, China. *BMC Microbiol*. 2015;15(1):6. doi:10.1186/s12866-015-0350-7

62. Groves HE, Papenburg J, Mehta K, et al. The effect of the COVID-19 pandemic on influenza-related hospitalization, intensive care admission and mortality in children in

Canada: A population-based study. *Lancet Reg Health Am*. 2022;7:100132. doi:10.1016/j.lana.2021.100132

63. Doroshenko A, Lee N, MacDonald C, Zelyas N, Asadi L, Kanji JN. Decline of Influenza and Respiratory Viruses With COVID-19 Public Health Measures: Alberta, Canada. *Mayo Clin Proc*. 2021;96(12):3042-3052. doi:10.1016/j.mayocp.2021.09.004

64. Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat Rev Genet*. 2016;17(5):257-271. doi:10.1038/nrg.2016.10

65. Rajagopala SV, Bakhoum NG, Pakala SB, et al. Metatranscriptomics to characterize respiratory virome, microbiome, and host response directly from clinical samples. *Cell Rep Methods*. 2021;1(6):100091. doi:10.1016/j.crmeth.2021.100091

66. Ali M, Han S, Gunst CJ, Lim S, Luinstra K, Smieja M. Throat and nasal swabs for molecular detection of respiratory viruses in acute pharyngitis. *Virol J*. 2015;12:178. doi:10.1186/s12985-015-0408-z

67. Goodall EC, Granados AC, Luinstra K, et al. Vitamin D3 and gargling for the prevention of upper respiratory tract infections: a randomized controlled trial. *BMC Infect Dis*. 2014;14:273. doi:10.1186/1471-2334-14-273

68. Silkoff PE, Flavin S, Gordon R, et al. Toll-like receptor 3 blockade in rhinovirus-induced experimental asthma exacerbations: A randomized controlled study. *J Allergy Clin Immunol*. 2018;141(4):1220-1230. doi:10.1016/j.jaci.2017.06.027

69. Cheemarla NR, Watkins TA, Mihaylova VT, et al. Dynamic innate immune response determines susceptibility to SARS-CoV-2 infection and early replication kinetics. *J Exp Med*. 2021;218(8):e20210583. doi:10.1084/jem.20210583

70. Chen LYC, Hoiland RL, Stukas S, Wellington CL, Sekhon MS. Confronting the controversy: interleukin-6 and the COVID-19 cytokine storm syndrome. *Eur Respir J*. 2020;56(4):2003006. doi:10.1183/13993003.03006-2020

71. Wang G, Wu C, Zhang Q, et al. C-Reactive Protein Level May Predict the Risk of COVID-19 Aggravation. *Open Forum Infect Dis*. 2020;7(5):ofaa153. doi:10.1093/ofid/ofaa153

72. Yang Y, Shen C, Li J, et al. Plasma IP-10 and MCP-3 levels are highly associated with disease severity and predict the progression of COVID-19. *J Allergy Clin Immunol*. 2020;146(1):119-127.e4. doi:10.1016/j.jaci.2020.04.027

73. Laing AG, Lorenc A, del Molino del Barrio I, et al. A dynamic COVID-19 immune signature includes associations with poor prognosis. *Nat Med*. 2020;26(10):1623-1635. doi:10.1038/s41591-020-1038-6

74. Mina MJ, Peto TE, García-Fiñana M, Semple MG, Buchan IE. Clarifying the evidence on SARS-CoV-2 antigen rapid tests in public health responses to COVID-19. *Lancet Lond Engl*. 2021;397(10283):1425-1427. doi:10.1016/S0140-6736(21)00425-6

75. Vogels CBF, Watkins AE, Harden CA, et al. SalivaDirect: A simplified and flexible platform to enhance SARS-CoV-2 testing capacity. *Med N Y N*. 2021;2(3):263-280.e6. doi:10.1016/j.medj.2020.12.010

76. Chaemsaithong P, Romero R, Korzeniewski SJ, et al. A point of care test for the determination of amniotic fluid interleukin-6 and the chemokine CXCL-10/IP-10. *J Matern-Fetal Neonatal Med Off J Eur Assoc Perinat Med Fed Asia Ocean Perinat Soc Int Soc Perinat Obstet*. 2015;28(13):1510-1519. doi:10.3109/14767058.2014.961417

77. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*. 2021;49(D1):D480-D489. doi:10.1093/nar/gkaa1100

# Chapter 3: Discussion and Future Directions:

## 3.1 Thesis Manuscript Limitations and Strengths

Historically, respiratory tract infections were common causes of socioeconomic stagnation. Pathogens such as the Spanish Flu, pneumonia, and tuberculosis had devastating and prolonged effects on populations throughout the world.[1,2] These infections, and many others like them, were responsible for the deaths of millions of people, recessions, and a depreciated standard of living. While significant improvements in modern medicine and the discovery of penicillin and other antibiotics have relieved the effects of these diseases, the recent impact of the COVID-19 pandemic has exposed major shortcomings in our collective understanding and ability to treat viral respiratory infections. Even prior to the pandemic, respiratory tract infections were reported as the dominant clinician-reported reasons for visiting primary care globally.[3] The human body is equipped with defence mechanisms to prevent respiratory tract viral infections *structurally* through the epithelium barrier and *molecularly* through an antiviral defence system, including the induction of ISGs. ISGs are a crucial part of the viral defence system. Biomarkers of respiratory tract infections have historically focused on the etiological cause of infection, although much of the morbidity and mortality is driven by the host-pathological response. Thus, ISGs in mucosal sites were explored for their role in viral defence and potential application as a host-biomarker of viral infection.

Our aim was to determine host biomarkers indicative of viral respiratory tract infections amenable to LFA testing. Data mining was performed on in-house, and publicly available datasets from RSV, RV, influenza A and SARS-CoV-2 infected patient nasopharyngeal swab samples and compared to negative controls. Using a qualitative selection criteria that included elevated presence in at least one dataset with viral infection, a protein product that is secreted, and that had antibodies commercially available for detection, *CXCL10*, *CXCL11* and *TNFSF10* gene expression levels were assessed, and a correlation analysis was performed with infection severity and time course. *CXCL10* and *CXCL11* upregulation was positively correlated with RSV compared to control ($p = 0.016$, $p = 0.006$). No significant association was found with influenza A or RV for all three genes. *CXCL10/CXCL11/TNFSF10* upregulation was positively correlated with SARS-CoV-2 infection compared to control ($p < 0.001$). *CXCL10* expression correlated with COVID-19 severity. CXCL10 was chosen as a lead biomarker candidate based on these multiple analyses that included different virus infections, time-courses, and measures of severity. CXCL10 was subsequently validated at the protein level in saliva as a prerequisite for developing a host-response LFA amenable to non-invasive, self-collection of samples by a non-expert. CXCL10 was not detected at the protein level in healthy saliva but was elevated in saliva from COVID-19 patients. A CXCL10 LFA was developed with a sensitivity of 2 ng/ml in a buffer and artificial saliva.

Interdisciplinary research is required to progress from patient engagement, clinical sample collection, pathogen diagnosis, transcriptomics, biomarker identification, and point-of-care diagnostic development. For the research in this thesis to be completed, engagement from consented patients, clinicians, clinical research coordinators, biomedical researchers, bioinformaticians, software developers, and industry partners was required. The interdisciplinary nature enables a breadth of study that cannot be achieved in isolation. However, this must be

balanced by the challenges of ensuring adequate depth in each discipline. The collaborative and integrative nature of the participants involved in this research included individuals from two hospitals, Sunnybrook Health Sciences Centre and at St. Joseph's Healthcare, two university labs, Hirota Lab at McMaster University and Doxey Lab at University of Waterloo, and a biotechnology company, Cytodiagnostics, for the co-development of saliva CXCL10 LFA. Beyond our direct resources for the project, indirect resources were procured from public databases and included large open-science datasets from the COVID-19 pandemic that enabled sample sizes (N =718 SARS-CoV-2 RNA-seq NPS) that would not have been achievable in isolation. Lastly, to provide the market-level assessment of any R&D activities in the lab, primary and secondary market research was performed via the Ryerson-Mitacs Lab2Market program to inform the end-to-end approach and context of clinical utility during early biomarker discovery research. Factors such as clinical utility, economics and the advantage of point-of-care testing platforms were explored. The arrangement and management of this constellation of resources posed both a challenge and an opportunity for the presented thesis, with the ultimate outcome demonstrating the ability for graduate studies to increase in both breadth (e.g., market analysis to prototype diagnostic) and depth (e.g., large sample size of well-phenotyped clinical samples) if non-traditional sources of data and partnerships are pursued.

Several limitations of our research must be mentioned for areas of future improvement and/or critical interpretation of the presented data. The first is the small sample sizes used for measuring gene expression from NPS in datasets that were curated from in-house and public resources. Specifically for the RSV, RV (n = 6 Control, n = 6 RSV, n = 12 RV) and FluA infected individuals (n = 8 FluA positive, n = 14 negative control), the sample sizes were small and may have influenced the ability to detect a signal in candidate ISG biomarkers. RNA is dynamic and prone to laboratory batch effects and can be compounded by intrinsic donor-to-donor variability in responses to respiratory tract infections [4]; therefore, a sample size of 6 may be unlikely to yield reproducible results.[5] A directly generated dataset (Sunnybrook Health Sciences Centre) assessing COVID-19 mortality was also a small sample size (n = 6 fatal, n = 19 not fatal) that limited the ability to confirm or refute a relationship with *CXCL10, CXCL11* and *TNFSF10* gene expression levels. To address this limitation, we took the indirect approach and curated a publicly available dataset[6] to analyze the relationship between viral load (Ct. threshold) and mortality with a much larger sample size of N = 735. Moving forward, a balance between using directly generating in-house data from well-phenotyped patients and publicly available datasets is likely to gain traction as the approach provides economic and time-efficiencies for independent confirmation of in-house data or testing of hypotheses. Integral to this occurring will be the quality of annotation of publicly available data and ease of access.

Small sample sizes were not limited to gene expression studies and were also a limitation for the CXCL10 analysis in the saliva of SARS-CoV-2 hospitalized COVID-19 patients. While the COVID-19 group showed a significantly greater CXCL10 concentration ($p < 0.05$), the data was asymmetrical. Two data points from the SARS-CoV-2 group in the figure appear to be far larger compared to the other four (see Figure 2.5B from thesis manuscript). Future studies should increase the sample size to determine if there are different clusters of COVID-19 patients in the context of CXLC10 levels and whether these clusters correlate with clinical outcomes. Unfortunately, the samples were collected during an early period of the COVID-19 pandemic prior to vaccinations, and it is unclear how samples collected post vaccination or from

individuals with prior history of SARS-CoV-2 infection will impact the CXCL10 levels. Simply increasing the sample size may address this limitation, but important metadata on patient vaccination status, variant the individual is infected with, and how many prior infections the individual has had, should also be collected in this post-COVID. Directly related to the CXCL10 detection in saliva via LFA, varying concentrations of CXCL10 through patient clusters (low CXCL10/COVID-19 group vs high CXCL10/COVID-19 group) presents a problem when using an LFA platform since it provides a binary result with a minimum limit of detection. Given the data distribution observed in Figure 2.5B, an optimal cut-off concentration of CXCL10 using a different technology platform may be a more appropriate testing platform.

Another limitation encountered during this research project was the unbalanced study demographics. The population in each analysis group was not matched for sex, age, smoking status, or race which could lead to confounding variables. For example, the RSV and RV[7] gene expression data was from children (3mo – 18yr), and SARS-CoV-2 gene expression data[6,8,9] was mainly from adults. There may be variables due to age differences that could affect expression levels of *CXCL10/CXCL11/TNFSF10*, separate from the focus of different viral infections. For example, in plasma samples, CXCL10 levels have been found to be significantly higher in older adults compared to younger adults.[10] In the future, with a stricter study design, study demographics could be more evenly matched to address the limitation.

Another limitation of the study was inconsistent technology platforms. Some gene expression data used microarray (Clariom™ D Assay), and others used RNA-sequencing technology (Illumina NextSeq or NovaSeq). Using incompatible technology platforms inhibits the ability to merge the datasets, which is why Figure 2.2 in the thesis manuscript separated each study into different figures, as it is not technically feasible to directly merge microarray and RNA-seq data into one analysis. RNA-seq is a more robust platform for differential gene expression analysis [11] due to the fact it has a broader range and can detect low abundance transcripts with higher sensitivity than microarray methods.[12] However, microarray technologies remain popular due to their cost and more straightforward analysis methods.[13]

Publicly available datasets provide opportunities for efficient secondary analysis by groups beyond the original authors, but this requires awareness that the data annotation, handling of specimens and technology platforms used may differ between studies. In the thesis manuscript, Figure 2.2 annotations and labelling of metadata slightly varied between control groups. For example, Mick et al[8] separated the SARS-CoV-2 negative cohort into 1) SARS-CoV-2 negative but positive for another respiratory virus and 2) SARS-CoV-2 negative with no other virus detected (non-viral ARIs). Other studies simply separated SARS-CoV-2 positive and negative. In addition, in the saliva SARS-CoV-2 CXCL10 analysis, the control group was healthy volunteers, which is not a clinically relevant study population. It is unlikely to encounter healthy people seeking medical care, more likely only those with a suspected respiratory infection. Lastly, there was variable quality of samples as there was no uniform procedure for handling, storage, technology platforms and analysis. Together these limitations can lead to bias, confounding factors, and inconstancies in the data.

Looking ahead at future directions for this project, more experimental research is required to further verify and validate the utility of CXCL10 as a biomarker of respiratory tract viral

infections. Specifically, correlating saliva CXCL10 with COVID-19 severity would require a prospective study with time-series measurements in well-phenotyped cases and controls, with a focus on in-patients as these are the most at risk for death. To determine if any saliva CXCL10 levels were specifically associated with COVID-19 severity, controls should include non-infected individuals, and hospitalized individuals infected with influenza or other acute non-infectious cases (e.g., trauma). The key for this study would be the detailed collection of study subject metadata that may track with saliva CXCL10 levels (e.g., vaccination status, variant infected with, number of infections). With the limited resources, funding, and budget available, this project attempted to view biomarker discovery from an end-to-end approach. Through utilizing the unique opportunity of SARS-CoV-2, it harnessed the wide range of publicly available datasets in addition to in-house datasets provided by local hospitals. The project also used a collaborative approach by partnering with another university laboratory and a biotechnology company. Our opinion is that this collaborative and interdisciplinary approach should continue to book-end the required prospective clinical study.

### 3.2 Path to Validation of Host Biomarker

This study observed several parameters, including *in-vitro* gene expression in relation to viral infection, viral load and SARS-CoV-2 mortality, and hospitalization. In addition, proteins found in healthy saliva and CXCL10 protein concentration in healthy and COVID-19-infected individuals were observed. Lastly, it was determined if the biomarker CXCL10 is amenable to LFA testing. The National Biomarker Alliance states in their best practices guidelines to clearly define the clinical question and that the sample type and context must be studied in context to the claimed clinical utility.[14] This project was exploratory and investigated several questions, but moving forward, a more defined clinical question must be stated.

*Is CXCL10 correlated with a viral infection? What virus? Compared to what control group?*
*Is CXCL10 prognostic of SARS-CoV-2 severity? What endpoint measures severity?*

The following steps to further develop this biomarker include analytical and clinical validation. Preferrable in a prospective blinded multi-centre study using sampling and analytical guidelines with defined guidelines and standard operating procedures (SOPs). Most of all, does the clinical question attempt to address an unmet need?[15] Can it help physicians answer questions and lead to improved decision-making? These items will be discussed in further detail (see **Table 3.1 - Case Study: IL-6 as a prognostic biomarker for COVID-19**), and how host biomarkers have the power to change the standard of care in infectious diseases with applications to discriminate viral and bacterial infections and the prognosis of infection.

The typical steps to validation in a biomarker development pipeline are shown in Figure 3.1. Several items should be considered during the study design: what disease, how many patients, matched controls, the sample type (blood, saliva, etc.), the class of molecule (mRNA, protein etc.), presence or absence (i.e., protein concentration cut off or positive/negative), profile of several biomarkers or just one (**Figure 3.3**).[16] Roche's Elecsys IL-6 assay assists in identifying severe inflammatory responses to COVID-19. The path to validation is explained and could be used as a model for a prognostic host biomarker for COVID-19.

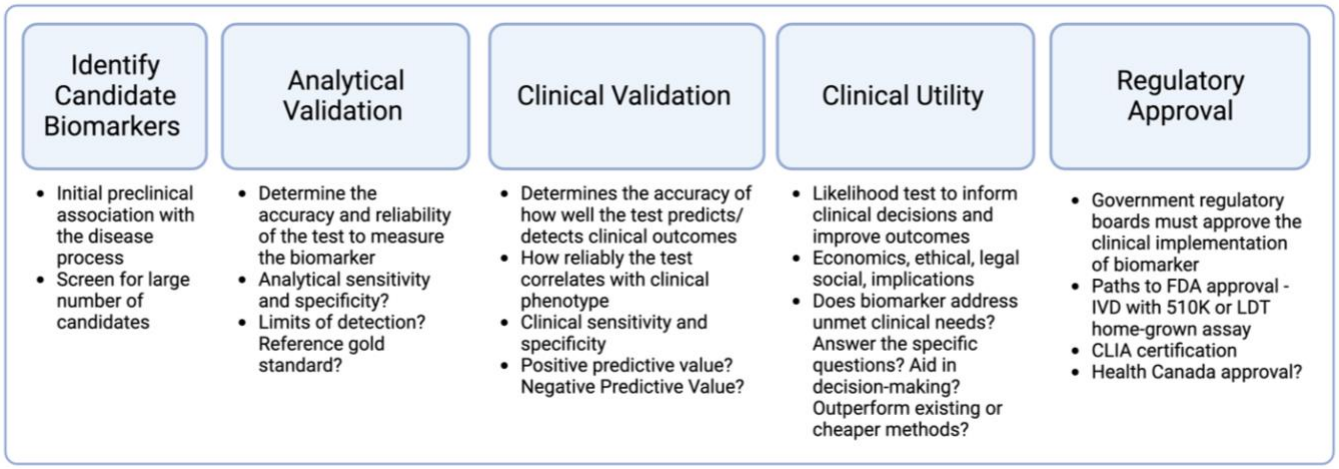



**Figure 3.1** Biomarker Development Pipeline. Adapted from Bime et al[17] *"The acute respiratory distress syndrome biomarker pipeline: crippling gaps between discovery and clinical utility"*& Byron et al[18] *"Translating RNA sequencing into clinical diagnostics: opportunities and challenges."*

1. *Biomarker Discovery*

During this phase, samples are screened for candidate biomarkers for an association with the disease [17]. Thousands of candidate biomarkers may be considered. Within host transcriptome studies, differential expression analysis is used to identify signatures of differentially expressed genes across different conditions or disease states.[19] Visualization of all genes can be performed with a volcano plot (**Figure 3.2**), highlighting those genes passing thresholds for fold change and statistical significance. Genes that have large fold changes **and** statistical significance between experimental groups become high priority candidates of interest.

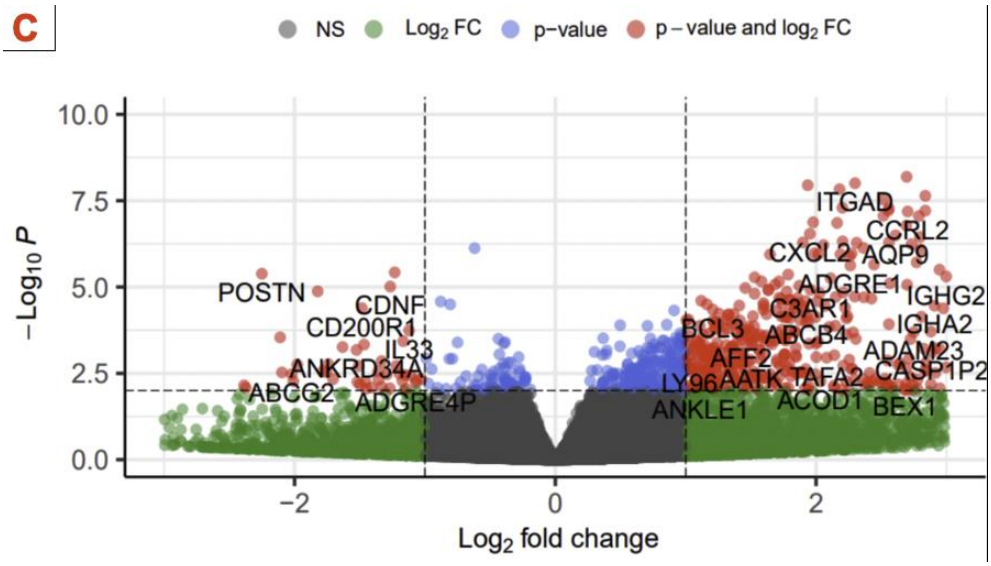



**Figure 3.2:** Volcano plot of RNA sequencing analysis of host transcriptomic responses in healthy control and patients with culture positive bacterial infected nasopharyngeal swab samples. Grey = non-significant, blue = statistically significantly different from healthy control ($-Log_{10}$ P>2). Green = $Log_2$ fold change (FC) relative to healthy control but not statistically significant, Red = $Log_2$ fold change relative to healthy control and statistically significant. *Unpublished data from Hirota/Doxey Labs.*

*2. Analytical and Clinical Validation*

The next phase in the biomarker development pipeline is analytical validation. The assay performance is determined in specific samples [17] here, one could test the sample using different platforms. Clinical validation determines the ability of the assay to predict clinical phenotypes.[18] The levels of detection (sensitivity) are also defined with known standards (positive and negative) validated within the assay. Specificity may be explored by analyzing samples spiked with common household (e.g., mouthwash, nasal spray) and medical reagents (corticosteroids, NSAIDs) to ensure these do not interfere with assay performance. Independent cohorts from multiple research sites are often deployed in this component of biomarker validation, as inter-site and inter-user variability must be considered as part of analytical validation. Within these independent cohorts, clinical validation of the original observation made in the discovery cohort can be confirmed. A variety of metrics may be used to calculate the clinical validity of diagnostic tests, including the positive and negative predictive values. A positive predictive value (PPV) determines the probability subjects with a positive test will have the disease/outcome, and a negative predictive value (NPV) determines the probability that subjects with a negative test will **not** have the disease/outcome.[18]

*3. Clinical Utility*

After the performance of the assay is assessed for how well is can identify a biomarker in a sample, and the associated phenotype, clinical utility is assessed. This step factors in the existing clinical workflow, economics, social and ethical considerations. How does the biomarker assay ultimately improve clinical outcomes and inform better decisions for health care practitioners and guide treatment options.[17,18]

*4. Regulatory Qualification and Approval*

The last but typically most challenging step is gaining Regulatory Qualification and Approval. There are two main paths to achieve regulatory approval with the FDA for a diagnostic – an in-vitro diagnostic with 510(k) or a laboratory-developed test (LDT).[17] As mentioned in the introduction, the average in-vitro diagnostic device costs $34 M and takes 6 years reach market approval.[20,21] With increasing inflationary pressures since this report, costs have continued to increase and show no signs of plateauing for receiving regulatory approval. The COVID-19 pandemic was an unprecedented time for regulatory qualification and approval. Most SARS-CoV-2 diagnostics were approved in record time (> 2 years) and granted emergency use authorization, including several SARS-CoV-2 rapid antigen tests[22] and Roche's Elecsys IL-6 IVD assay.[23]

**Table 3.1** Case Study: IL-6 as a prognostic biomarker for COVID-19. IL-6 as a predictive biomarker for COVID-19 death[24] and risk of intubation.[25]

**Case Study: IL-6 as a prognostic biomarker for COVID-19**

An example of a clinically validated and approved prognostic biomarker of viral infection is the cytokine IL-6. Serum IL-6 levels were found to be predictive of severe COVID-19 and the associated cytokine storm.[26] Roche's Elecsys IL-6 electrochemiluminescence immunoassay rapidly identifies serum of plasma IL-6 concentrations.[23] The two studies below examine circulating IL-6 levels in COVID-19 patients. Its path to validation can serve as a potential template for designing other studies of prognostic biomarkers of viral infection severity.

**IL-6 as a predictive biomarker for COVID-19 death: Zhang et al**[24]
This retrospective institutional study measured serum IL-6 during a hospital stay. Patients were stratified as common (n = 366), severe (n = 411), or critical (n =124) based on existing guidelines. Serum IL-6 concentrations predictive of death were calculated. They found that serum IL-6 concentrations higher than 37.65 pg/ml were predictive of death (sensitivity = 91.7%, specificity 95.7%). In addition, the study prescribed tocilizumab (IL-6 receptor blockade) to 16 patients and found no significant differences in disease outcomes compared to matched counterparts.

**IL-6 and of risk of intubation: Herold et al**[25]
IL-6 was validated in an initial(n=40) and validation (n=49) cohort to predict the need for mechanical ventilation during SARS-CoV-2 infection. Through this study, IL-6 concentrations from hospitalized SARS-Cov-2 patients were determined at assessment, and over their stay in the hospital, with the endpoint defined as respiratory failure requiring mechanical ventilation. IL-6 levels at assessment (>35 pg/ml) and maximal values over hospital stay (>80 pg/ml) predicted the need for intubation. Area under curve (AUC) values and optimal cut-off concentrations were calculated. In addition, the negative and positive predictive values were calculated. The study found Roche's Elecsys IL-6 demonstrated an effective aid in determining hospitalized patients at risk for mechanical ventilation. To demonstrate its clinical performance, 49 PCR-positive patients enrolled in the emergency room. Using the Elecsys IL-6 assay with a cut-off of IL-6 > 35pg/ml was able to identify 16/19 patients that required intubation.

The FDA approved Roche's Elecsys IL-6 to identify hospitalized COVID-19 patients' risk of intubation with mechanical ventilation from serum or plasma.[23] Roche acknowledged that this test should be used alongside other clinical parameters.[23]

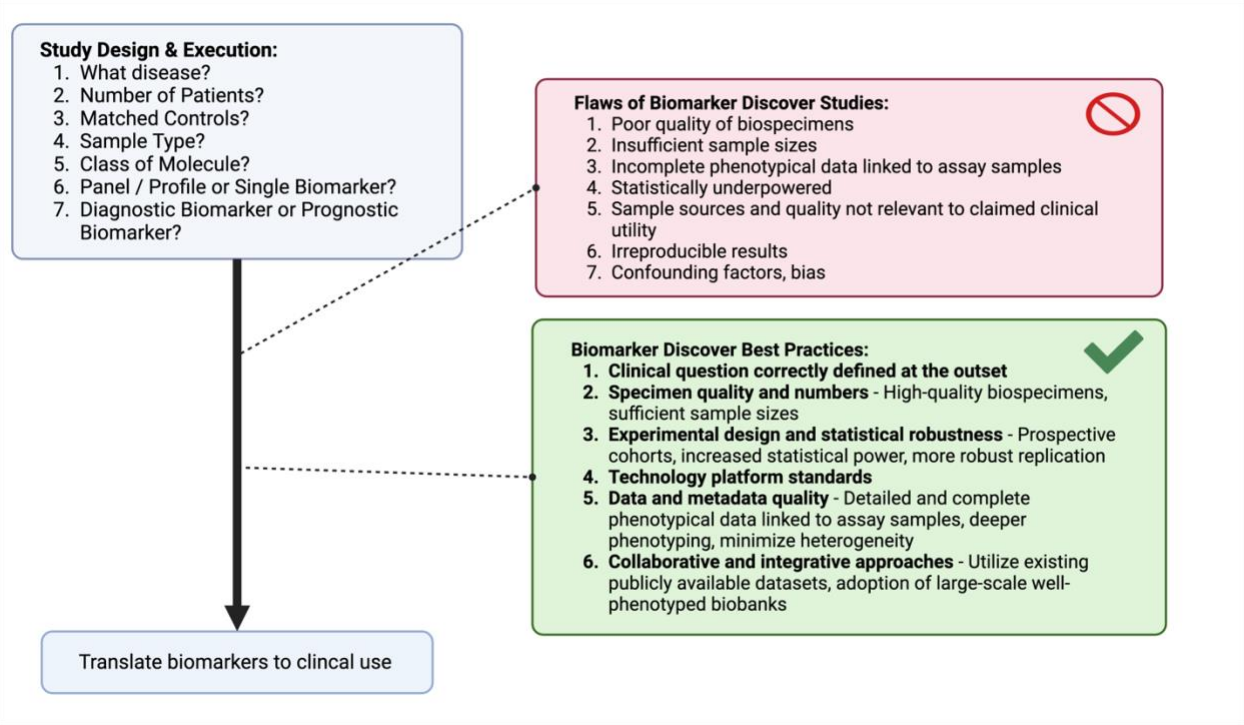



**Figure 3.3** Study Design Biomarker Development: best practices and flaws). *Adapted from:* Barker et al,[14] Drabovich et al,[15] Issaq and Veenstra[16] and Bime et al.[17]

## 3.3 Hypothetical Study to Validate CXCL10 as a Biomarker of Respiratory Tract Infection from Mucosal Sampling

*CXCL10 as a biomarker of viral infection (Figure 3.4A)*

Figure 3.4A outlines a study design to validate CXCL10 protein levels in saliva to discriminate between viral or bacterial infections. The clinical utility of discriminating between a viral and bacterial infection could aid in decision making to prescribe antibiotics. CXCL10 levels could be potentially used to identify a viral infection or suggest bacterial infection and the need for antibiotics.

The MeMed Opportunity study[27] used TRAIL, CRP and CXCL10 to discriminate between bacterial and virial respiratory tract infections. The double blind, multi-centre validation study tested the performance of the assay in children aged 2-60 months. They assessed 577 children with suspected acute respiratory infection (n = 71 bacterial, n = 435 viral). Since there is no universally gold standard to determine the cause of an acute respiratory infection, a diagnosis was made based off a panel of three pediatricians with all available laboratory and clinical data. The diagnosis was separated into unanimous agreement, majority or inconclusive by the results from the pediatricians. The results showed the TRAIL, CRP and CXCL10 assay was successfully able to identify bacterial infections from viral (sensitivity = 86.7%, specificity = 91.1%) and can aid in antimicrobial stewardship.[27]

Following a similar study design to the MeMed Opportunity study [27] patients would need to be enrolled suspect of acute respiratory infection, and diagnosed as bacterial or viral, confirmed by an expert panel of clinicians and by laboratory methods. Patients with inconclusive results would not be included. CXCL10 levels should be measured upon admission, and ROC curve could determine the optimal cut-off concentration to be used, suggestive of viral, bacterial, or unknown (Figure 3.4A). The performance of this test could be measured against current markers of inflammation, similar to the MeMed Opportunity study, which used C-reactive protein and procalcitonin. The main difference between our study, is using saliva instead of blood, and only a single biomarker.

*CXCL10 as a biomarker of disease severity (Figure 3.4B)*

A different study design would be required if one wanted to validate association between COVID-19 disease severity and CXCL10 saliva protein concentrations (Figure 3.4B). Patients who are hospitalized for COVID-19 would be enrolled with positive PCR tests. Patients would be stratified based on mild, moderate, severe, or critical illness based off the NIH guidelines.[28] CXCL10 concentration in saliva would be measured upon admission and subsequent days spent in the hospital to determine CXCL10 kinetics. Endpoints defined as discharged, continued hospitalization (within time frame of the study) and death; ROC curve could determine the optimal cut-off concentration to be used predictive of each COVID-19 outcome.


**Figure 3.4** Study Design to Validate CXCL10 levels in saliva correlation **A** Bacterial / Viral Infection **B** with SARS-COV-2 mortality. Definition of COVID-19 Severity by National Insitiue of Health.[28]

### 3.4 Host-biomarkers Changing the Standard of Care in Respiratory Infections

The applications of using a host in respiratory infections span greater than discriminating between viral and bacterial infections. There is potential to define molecular signatures of disease in heterogeneous syndromes such as ARDS.[17,29] This can assist in removing bias in clinical trials and identifying responders to certain drugs. In addition, host biomarkers can be used to develop new drug targets that can ameliorate immunopathology.[30] This was observed in Zhang study, which prescribed tocilizumab (IL-6 receptor blockade) to 16 patients and measured to IL-6 serum levels.[24] Most of these patients showed increased in IL-6 levels after receiving tocilizumab but no improvements in survival outcomes was observed.[24] In addition, there may be anti-inflammatory cytokines and molecules which help to augment immunity.[30] Lastly, host biomarkers can be applied as a prognostic biomarker and allow for early intervention.

*The barrier to clinical implementation*

As discussed in the introduction, the current state of medical care is decades away from implementing host biomarkers into point of need diagnostics performed by non-healthcare professionals. The fraction of biomarker candidates that reach regulatory approval is very slim. In cancer research, less than 1% of candidate biomarkers published successfully enter clinical practice.[31] Referred to as the "valley of death," [32] this discrepancy between researched biomarker candidates and approved clinical biomarkers is real. Most biomarker discovery occurs within university laboratories similar to this project, and few have the resources to continue clinical validation and achieve regulatory approval.[14] A wealth of papers within the acute respiratory infection space attempt to answer the same question and identify a molecular signature indicative of viral or bacterial infection. Perhaps different classes of molecules and different sample types, still, they were all looking for panels that discriminated between viral and bacterial infection in respiratory infections.[27,33–35]

In light of open science and the open sharing of data, it may be valuable to generate a database for host response biomarkers to infections. Researchers could deposit data from their study provided uniform specimens, technology platforms and terminology are used. Through this study, we attempted to use existing databases such as NCBI GEO. However, limitations of different study designs, the definition of control groups and specimen quality prevented the ability to merge all findings together. However, if national and international scientific communities (Like the National Biomarker Alliance) worked together to build a reliable database, it would require guidelines and SOPs to ensure uniform sample handling and technology platforms are used. The CARD database is an example of a database that this could emulate. Perhaps, an ontology with a clear definition of relationships would help to address ad-hoc data and variability between studies. A database may also help to address the issue of insufficient sample sizes, prevent researchers from repeating the same experiments and allow smaller university laboratories to access big data. Ultimately make the best use of investments, resources, and patient samples.[14]

*Success stories and looking to the future*

Through the timespan of this thesis, host-based diagnostics have made significant progress and gained further attention. Since January 2021, when this project began, millions of dollars have been funded companies in the space of host-based diagnostics for infectious diseases, including Inflammatix and MeMed. These host-biomarkers assays are valuable economically and can improve patient care.

As mentioned above, MeMed BV® measures TRAIL, IP-10, and CRP in serum samples to differentiate bacterial from viral infection in 15 minutes.[36] Progressing from the OPPORUNITY study[27], a clinical trial was conducted in 2020 to determine the diagnostic performance of MeMed BV®. The test's performance was determined in pediatric and adult patients (n=1384) in MeMed's Apollo clinical trial (NCT04690569). The MeMed BV® IVD test (510(k): K210254) gained CE mark at the June 2020[37] and FDA approval in September 2021.[38] As of 2022, Maimonides Medical Center in Brooklyn, NY and Maccabi Healthcare Services in Israel announced they are integrating MeMed BV® as part of routine care.[39,40]

> *"The new test [MeMed BV®] is a groundbreaking technology that allows for a faster and more accurate diagnosis in community medicine. The test is intended for use in cases where there is uncertainty around the origin of an infection – whether it is bacterial or viral – and the usual auxiliary tests are not sufficient. The test is especially beneficial in children and elderly patients who are at higher risk of medical deterioration if a bacterial infection is not treated in time. We have no doubt that this test can save lives."*
> - *Dr. Miri Mizrahi Reuven (Head of Maccabi's health division)*[40]

Aside from respiratory infections, another company, Inflammatix, has been working on TriVerity, a host-biomarker test to identify bacterial or viral causes in Sepsis patients in the ICU, aiming to identify the "presence, type and severity" of infection. The TriVerity assay is done using blood samples, with an mRNA panel of 29 genes, to determine the cause of infection (bacterial or viral) and disease outcome.[41] The company received $102 million funding for its Series D funding in March 2021.[42] However, none of their tests have reached FDA clearance.

Collectively, these success stories demonstrate the potential value that lies within host biomarkers being used in infectious disease diagnostics. The pathway to validation appears to be a long and expensive journey. However, these will lead to improved patient care. As said previously, most biomarkers research will never reach regulatory approval. Hopefully, MeMed's journey and study designs can pave the way for other host biomarkers where the current practices and tests fail to serve patients and healthcare practitioners.

## 3.5 Conclusion

Our goal was to determine host biomarkers indicative of viral respiratory tract infections. This thesis strived to take advantage of the wealth of open-access research published during the COVID-19 pandemic utilizing local partnerships and publicly available research. We explored ISGs for their potential as a biomarker from mucosal samples and found significant results for CXCL10 as a biomarker of viral infection. A prototype point of care CXCL10 responsive LFA was developed. By researching other host biomarkers and conducting market research through the L2M program, validating these biomarkers and reaching regulatory clearance is a complicated journey requiring compressive large-scale study designs and clinical trials.

### 3.6 References:


1. Morens DM, Taubenberger JK, Fauci AS. Predominant Role of Bacterial Pneumonia as a Cause of Death in Pandemic Influenza: Implications for Pandemic Influenza Preparedness. *J Infect Dis*. 2008;198(7):962-970. doi:10.1086/591708

2. Pai M, Behr MA, Dowdy D, et al. Tuberculosis. *Nat Rev Dis Primer*. 2016;2(1):1-23. doi:10.1038/nrdp.2016.76

3. Finley CR, Chan DS, Ccfp SG, et al. What are the most common conditions in primary care? :9.

4. Goh WWB, Wang W, Wong L. Why Batch Effects Matter in Omics Data, and How to Avoid Them. *Trends Biotechnol*. 2017;35(6):498-507. doi:10.1016/j.tibtech.2017.02.012

5. Maleki F, Ovens K, McQuillan I, Kusalik AJ. Size matters: how sample size affects the reproducibility and specificity of gene set analysis. *Hum Genomics*. 2019;13(1):42. doi:10.1186/s40246-019-0226-2

6. Butler D, Mozsary C, Meydan C, et al. Shotgun transcriptome, spatial omics, and isothermal profiling of SARS-CoV-2 infection reveals unique host responses, viral diversification, and drug interactions. *Nat Commun*. 2021;12(1):1660. doi:10.1038/s41467-021-21361-7

7. Yu J, Peterson DR, Baran AM, et al. Host Gene Expression in Nose and Blood for the Diagnosis of Viral Respiratory Infection. *J Infect Dis*. 2019;219(7):1151-1161. doi:10.1093/infdis/jiy608

8. Mick E, Kamm J, Pisco AO, et al. Upper airway gene expression reveals suppressed immune responses to SARS-CoV-2 compared with other respiratory viruses. *Nat Commun*. 2020;11(1):5854. doi:10.1038/s41467-020-19587-y

9. Lieberman NAP, Peddu V, Xie H, et al. In vivo antiviral host transcriptional response to SARS-CoV-2 by viral load, sex, and age. Cadwell K, ed. *PLOS Biol*. 2020;18(9):e3000849. doi:10.1371/journal.pbio.3000849

10. Bradburn S, McPhee J, Bagley L, et al. Dysregulation of C-X-C motif ligand 10 during aging and association with cognitive performance. *Neurobiol Aging*. 2018;63:54-64. doi:10.1016/j.neurobiolaging.2017.11.009

11. Li J, Hou R, Niu X, et al. Comparison of microarray and RNA-Seq analysis of mRNA expression in dermal mesenchymal stem cells. *Biotechnol Lett*. 2016;38(1):33-41. doi:10.1007/s10529-015-1963-5

12. Rao MS, Van Vleet TR, Ciurlionis R, et al. Comparison of RNA-Seq and Microarray Gene Expression Platforms for the Toxicogenomic Evaluation of Liver From Short-Term Rat Toxicity Studies. *Front Genet*. 2019;9:636. doi:10.3389/fgene.2018.00636

13. Rai MF, Tycksen ED, Sandell LJ, Brophy RH. Advantages of RNA-seq Compared to RNA Microarrays for Transcriptome Profiling of Anterior Cruciate Ligament Tears. *J Orthop Res Off Publ Orthop Res Soc*. 2018;36(1):484-497. doi:10.1002/jor.23661

14. Barker A, Compton C, Poste G. The National Biomarker Development Alliance accelerating the translation of biomarkers to the clinic. *Biomark Med*. 2014;8:873-876. doi:10.2217/bmm.14.52

15. Drabovich AP, Pavlou MP, Batruch I, Diamandis EP. Chapter 2 - Proteomic and mass spectrometry technologies for biomarker discovery. In: Issaq HJ, Veenstra TD, eds. *Proteomic and Metabolomic Approaches to Biomarker Discovery (Second Edition)*. Academic Press; 2013:17-37. doi:10.1016/B978-0-12-818607-7.00002-5

16. Issaq HJ, Veenstra TD. Chapter 1 - Biomarker discovery: Study design and execution. In: Issaq HJ, Veenstra TD, eds. *Proteomic and Metabolomic Approaches to Biomarker Discovery (Second Edition)*. Academic Press; 2020:1-16. doi:10.1016/B978-0-12-818607-7.00001-3

17. Bime C, Camp SM, Casanova N, et al. The acute respiratory distress syndrome biomarker pipeline: crippling gaps between discovery and clinical utility. *Transl Res*. 2020;226:105-115. doi:10.1016/j.trsl.2020.06.010

18. Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat Rev Genet*. 2016;17(5):257-271. doi:10.1038/nrg.2016.10

19. McDermaid A, Monier B, Zhao J, Liu B, Ma Q. Interpretation of differential gene expression results of RNA-seq data: review and integration. *Brief Bioinform*. 2018;20(6):2044-2054. doi:10.1093/bib/bby067

20. Chin CD, Linder V, Sia SK. Commercialization of microfluidic point-of-care diagnostic devices. *Lab Chip*. 2012;12(12):2118. doi:10.1039/c2lc21204h

21. Drlik M. How Much Does it Cost to Develop a Medical Device? FOUR METHODS TO ESTIMATE MEDICAL DEVICE AND POINT OF CARE DIAGNOSTIC DEVELOPMENT COSTS.

22. Health Canada. COVID-19 testing devices: Authorized medical devices. Government of Canada. Published September 13, 2020. Accessed October 20, 2022. https://www.canada.ca/en/health-canada/services/drugs-health-products/covid19-industry/medical-devices/authorized/list.html

23. FDA. Elecsys IL-6 Package Insert 2020-06, V1.0; Material Numbers 05109442190 and 07027532190.

24. Zhang J, Hao Y, Ou W, et al. Serum interleukin-6 is an indicator for severity in 901 patients with SARS-CoV-2 infection: a cohort study. *J Transl Med*. 2020;18(1):406. doi:10.1186/s12967-020-02571-x

25. Herold T, Jurinovic V, Arnreich C, et al. Elevated levels of IL-6 and CRP predict the need for mechanical ventilation in COVID-19. *J Allergy Clin Immunol*. 2020;146(1):128-136.e4. doi:10.1016/j.jaci.2020.05.008

26. Del Valle DM, Kim-Schulze S, Huang HH, et al. An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nat Med*. 2020;26(10):1636-1643. doi:10.1038/s41591-020-1051-9

27. van Houten CB, de Groot JAH, Klein A, et al. A host-protein based assay to differentiate between bacterial and viral infections in preschool children (OPPORTUNITY): a double-blind, multicentre, validation study. *Lancet Infect Dis*. 2017;17(4):431-440. doi:10.1016/S1473-3099(16)30519-9

28. NIH. National Institutes of Health's Coronavirus Disease (COVID-19) Treatment Guidelines. Published online September 26, 2022. https://www.covid19treatmentguidelines.nih.gov/.

29. Maslove DM, Tang B, Shankar-Hari M, et al. Redefining critical illness. *Nat Med*. 2022;28(6):1141-1148. doi:10.1038/s41591-022-01843-x

30. Wallis RS, O'Garra A, Sher A, Wack A. Host-directed immunotherapy of viral and bacterial infections: past, present and future. *Nat Rev Immunol*. Published online June 7, 2022:1-13. doi:10.1038/s41577-022-00734-z

31. Kern SE. Why your new cancer biomarker may never work: Recurrent patterns and remarkable diversity in biomarker failures. *Cancer Res*. 2012;72(23):6097-6101. doi:10.1158/0008-5472.CAN-12-3232

32. Seyhan AA. Lost in translation: the valley of death across preclinical and clinical divide – identification of problems and overcoming obstacles. *Transl Med Commun*. 2019;4(1):18. doi:10.1186/s41231-019-0050-7

33. Bhattacharya S, Rosenberg AF, Peterson DR, et al. Transcriptomic Biomarkers to Discriminate Bacterial from Nonbacterial Infection in Adults Hospitalized with Respiratory Illness. *Sci Rep*. 2017;7(1):6548. doi:10.1038/s41598-017-06738-3

34. Langelier C, Kalantar KL, Moazed F, et al. Integrating host response and unbiased microbe detection for lower respiratory tract infection diagnosis in critically ill adults. *Proc Natl Acad Sci U S A*. 2018;115(52):E12353-E12362. doi:10.1073/pnas.1809700115

35. Ross M, Henao R, Burke TW, et al. A comparison of host response strategies to distinguish bacterial and viral infection. *PLoS ONE*. 2021;16(12):e0261385. doi:10.1371/journal.pone.0261385

36. Ashkenazi-Hoffnung L, Oved K, Navon R, et al. A host-protein signature is superior to other biomarkers for differentiating between bacterial and viral disease in patients with respiratory infection and fever without source: a prospective observational study. *Eur J Clin Microbiol*

*Infect Dis Off Publ Eur Soc Clin Microbiol*. 2018;37(7):1361-1371. doi:10.1007/s10096-018-3261-3

37. MeMed. MeMed Receives CE Mark for Two Pioneering Technologies: The MeMed BV™ Test and MeMed Key™ Point-of-Need Platform. GlobeNewswire News Room. Published February 6, 2020. Accessed October 18, 2022. https://www.globenewswire.com/en/news-release/2020/06/02/2041893/0/en/MeMed-Receives-CE-Mark-for-Two-Pioneering-Technologies-The-MeMed-BV-Test-and-MeMed-Key-Point-of-Need-Platform.html

38. US approves MeMed test to distinguish bacterial from viral infections. Medical Device Network. Published September 20, 2021. Accessed October 18, 2022. https://www.medicaldevice-network.com/news/memed-bacteria-virus-test/

39. Maimonides Medical Center Adopts Revolutionary Technology That Quickly Distinguishes Bacterial and Viral Infections. AP NEWS. Published March 22, 2022. Accessed October 18, 2022. https://apnews.com/press-release/globe-newswire/technology-health-new-york-brooklyn-medical-technology-47b3ad395aaf894ac4ad7a7439733aea

40. MeMed. Maccabi Healthcare Services announces it is the first HMO in the world to include MeMed BV® test as part of routine care. GlobeNewswire News Room. Published July 6, 2022. Accessed October 18, 2022. https://www.globenewswire.com/en/news-release/2022/06/07/2457291/0/en/Maccabi-Healthcare-Services-announces-it-is-the-first-HMO-in-the-world-to-include-MeMed-BV-test-as-part-of-routine-care.html

41. Brakenridge SC, Chen UI, Loftus T, et al. Evaluation of a Multivalent Transcriptomic Metric for Diagnosing Surgical Sepsis and Estimating Mortality Among Critically Ill Patients. *JAMA Netw Open*. 2022;5(7):e2221520. doi:10.1001/jamanetworkopen.2022.21520

42. Inflammatix Secures $102 Million in Series D Financing Led by D1 Capital Partners to Commercialize Novel Host Response. *Bloomberg.com*. https://www.bloomberg.com/press-releases/2021-03-16/inflammatix-secures-102-million-in-series-d-financing-led-by-d1-capital-partners-to-commercialize-novel-host-response. Published March 16, 2021. Accessed October 19, 2022.

# Appendices

| Study (GSE) | Number of Microarray Datasets | Cell Line | Virus |
|---|---|---|---|
| 980 | 2 | Human CD14+ monocytes | Measles |
| 3397 | 2 | Human bronchial epithelial cells (BEAS-2B cells) | RSV |
| 8961 | 5 | A549 cells | HMPV |
| 17400 | 3 | 2B4 | SARS-CoV, DHOV |
| 28166 | 6 | Calu-3 cells | H5N1 Avian Influenza Virus |
| 28904 | 1 | B2B and 16HBE | Rhinovirus |
| 32138 | 1 | airway epithelial cell cultures | FluA or RSV |
| 33267 | 22 | 2B4 cells | icSARS CoV or the icSARS deltaORF6 mutant. |
| 37571 | 9 | Calu-3 cells | A/CA/04/2009 Influenza |
| 37827 | 19 | 2B4 cells | icSARS CoV or the icSARS deltaORF6 mutant. |
| 40844 | 13 | Calu-3 cells | H1N1 influenza virus A/Netherlands/602/2009 |
| 43203 | 14 | Calu-3 | Influenza viruses |
| 43204 | 7 | Calu-3 cells | S1trunc124: A/Vietnam/1203-CIP048_RG4/2004 (H5N1) or WT: A/Vietnam/1203/2004 (H5N1) |
| 45042 | 4 | Calu-3 2B4 cells | Human Coronavirus EMC 2012 (HCoV-EMC) |
| 47960 | 32 | HAE cultures | SARS-CoV, SARS-dORF6 or SARS-BatSRBD |
| 47961 | 26 | HAE cultures. | SARS-CoV, SARS-dORF6 and SARS-BatSRBD |
| 47962 | 33 | HAE cultures. | SARS-CoV, SARS-dORF6 and SARS-BatSRBD |

**Supplementary Table 1**: Data from Harmonizome used in Figure 1. Consisted of 199 datasets came from 17 GSE studies including the following viruses Influenza A, Human coronavirus, SARS-CoV like virus isolated from bats, A mouse-adapted SARS-coronavirus, SARS-CoV mutant strain that does not express the accessory protein open reading frame 6 (delta ORF6), Human metapneumovirus, Wild-type infectious clone-derived SARS-CoV, Measles, RSV and Rhinovirus and time points ranging from 0- 96 hours post inoculation. *In-vitro* cell systems included but were not limited to Calu-3, HAE, and A549 cell lines