Responding to Emerging Infectious Diseases with Big Data, Rapid Diagnostics, and Stem Cell Technology

by

Qing Yang

B.S., Microbiology, University of Minnesota Twin Cities, 2017

B.A., Computer Science, University of Minnesota Twin Cities, 2017

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Molecular, Cellular and Developmental Biology

2022

Committee Members:

Edward B. Chuong, PhD, Chair

Sara L. Sawyer, PhD

Robin D. Dowell, ScD

Mary A. Allen, PhD

Roy R. Parker, PhD

Yang, Qing (Ph.D., Molecular, Cellular and Developmental Biology)

Responding to Emerging Infectious Diseases with Big Data, Rapid Diagnostics, and Stem Cell Technology

Thesis directed by Prof. Sara L. Sawyer, PhD and Prof. Robin D. Dowell, ScD

The ongoing COVID-19 pandemic highlights the need for adapting high-throughput methods to identify, characterize, and prevent infectious diseases. Prior to the pandemic, my dissertation focused on detecting infections using host immune transcriptional signatures. To detect diverse infections, independent of the pathogen-specific molecules, I focused on identifying common host RNA transcripts upregulated in a large collection of transcriptomic datasets. The identified panel of host RNA transcripts was validated as detectable from *in vitro* models to the blood, serum, and saliva of infected individuals. Consistent with our understanding of host innate immunity, these transcripts are largely derived from genes related to type-I interferon and inflammatory responses. I then validated these findings using saliva samples from SARS-CoV-2 infected individuals and showed that this panel of host RNA transcripts can reliably identify infected individuals independent of their symptomatic status. Together, I showed that specific host RNAs, present in saliva, can serve as reliable markers for detecting ongoing infection, independent of pathogen abundance and disease-associated symptoms.

In response to the current pandemic, I worked with my colleagues during the campus shutdown on the development and deployment of a rapid saliva-based SARS-CoV-2 nucleic acid screening test to be used on our university campus. The rapid screening test required only a simple set of pipettes and a heating block, and effectively bypassed the supply chain issues that were slowing down testing efforts nationwide at that time. Eventually, the test was deployed all over the state of Colorado and in other countries in need. Through our campus screening efforts, we also identified over 1,500 saliva samples from asymptomatic

SARS-CoV-2 infected individuals. This rare collection of samples allowed us to compare symptomatic to asymptomatic infections. Interestingly, I found that both populations carry equivalent viral loads, and that only 2% of SARS-CoV-2 infected individuals harbored over 90% of the viral genomes circulating in the community. This further provided the first evidence for viral supercarriers (individuals carrying orders of magnitude more virus than the average).

Ultimately, I want the world to be better prepared for the next emerging infectious disease. During the study of host response to viral infection, I realized the caveats of using immortalized cell lines as infection models, as they often fail to fully represent the true scale of human immune response. I thus sought to validate an induced pluripotent stem cell (iPSCs)-derived monocyte/macrophage system. Monocytes and monocyte-derived macrophages function as key mediators of inflammation and antiviral immunity. At the same time, they are one of the key target cell types for numerous viruses with pandemic potential (e.g. arterivirus and Ebola virus). However, macrophages have relatively low abundance in the bloodstream and cannot be propagated ex vivo, meaning that consistent access to blood donors is required for routine research. To overcome these hurdles, I differentiated monocyte-derived macrophages from iPSCs as an alternative model system. Since iPSCs proliferate indefinitely and can be expanded to large quantities, I was able to isolate millions of iPSC-derived macrophages over a short period. Using phenotypic profiling of surface marker expression, transcriptomics (RNA-seq), and genome accessibility (ATAC-seq) profiling, I showed that iPSC-derived macrophages are nearly identical to their blood-derived counterparts. Furthermore, I showed that iPSC-derived macrophages support the replication of viruses, including HIV-1, dengue virus, and influenza virus, to the same degree as the primary macrophages. Our results validate iPSC-derived macrophages as a valuable model system for studying host-virus interactions.

Dedication

To all who seek truth and push science forward with intelligence and dedication.

And to my mentors, friends and family.

Acknowledgements

The most important lesson I learned from my graduate school experience is that no science discovery is made by oneself alone. I am grateful that I got to experience it alongside so many brilliant minds in this highly collaborative and supportive environment. I thank my mentors, Sara Sawyer, Robin Dowell, Mary Allen, Ed Chuong and Roy Parker. You inspire and guide me with your expertise, ethics and intelligence. Specifically, I am so fortunate that I got to be co-advised by Sara and Robin. You set the best examples of a dedicated and rigorous scientist. And the skills you taught me in problem solving, scientific writing and presentation will benefit me for life. Moreover, a lab wouldn't be complete without the hardworking trainees. I am grateful to Nick Meyerson, Maryska Kaczmarek and Jacob Stanley for welcoming me into the labs and getting me up to speed as a fresh graduate student with countless naive questions. It was Nick Meyerson, Cody Warren, Alison Gilchrist, Daniel Ramirez, Camille Paige, Will Fattor, Kyle Clark and Arturo Barbachano-Guerrero who offered me their hands (and beers) to help and support me through many failed experiments, stressful presentations, and mountains of saliva samples. And I will forever miss every past and current members of Sawyer, Dowell and Allen labs along with all my graduate school friends, thank you for creating such a loving, creative and supportive environment for me to grow.

I am also so fortunate to have the best family I could hope for. Especially, I have a mother who is both rigorous and nurturing - I was told to communicate through e-mails, use shared calendars and send her monthly progress reports since I was eight! But when

I switched major four times in college and when I decided to go to graduate school, all I heard from my mother was: "If you think it's the right move, go for it!". Last but not least, meeting, getting to know, and marrying my wife, Sydney Miller, during the graduate school was probably the best thing that has happened. Knowing that I can always come home to her helped me through the most stressful times. Through our epic weekend adventures, she brought so much happiness into my life.

Thank you all!

Contents

Chapter

Tables

Table

Figures

Figure

Chapter 1

Introduction

1.1 Preface: Response to emerging infectious diseases

Humans coexist and interact with a diverse range of pathogens. In a rare event, the transmission of the infectious pathogen occurs across multiple continents, affecting a large portion of the human population, resulting in a pandemic. In the past century, humans have encountered numerous infectious diseases that reached the pandemic level, including the ongoing human immunodeficiency virus (HIV) and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemics [1]. In response to a pandemic, efficient public health measures combined with a comprehensive understanding of the pathogen-host interaction can help mitigate the severity by controlling the disease transmission and offering potential treatment.

Our response to the ongoing coronavirus disease 2019 (COVID-19)/SARS-CoV-2 pandemic exemplifies the modern approach to counteract the infectious disease. Namely, the approach adapts technology advancements in the fields of communication, molecular biology and computer science: **(1)** High-throughput viral genome sequencing and data sharing enabled virus variant monitoring and a synchronized research effort globally; **(2)** Daily case reporting and epidemiological modeling enabled implementation of adaptive public health measures; **(3)** The mRNA-based vaccine technology accelerated the process of immunizing the susceptible population, and **(4)** Efficient drug screening provided antiviral treatments to lower the disease fatality rate.

However, the ongoing pandemic also exposes several fundamental shortcomings in the response: **(1)** A lack of understanding of the disease dynamic led to non-specific public health measures that sometimes failed to limit the transmission. Specifically, the asymptomatic transmission of SARS-CoV-2 is rarely accounted for, due to the lack of effective screening tools [2]; **(2)** Research designs that lacked biological relevance led to ineffective antiviral treatments. For example, the use of irrelevant *in vitro* models that failed to accurately capture the SARS-CoV-2 entry led to the controversial usage of chloroquine as a COVID-19 treatment [3, 4].

Realizing these shortcomings, while being actively engaged in the local effort of pandemic response, I aim to address some of these shortcomings through my dissertation by **(1) highlighting the prevalence of SARS-CoV-2 asymptomatic population, (2) providing efficient infection screening tools for detecting asymptomatic individuals, and (3) establishing more relevant model systems for future pathogen-host interaction studies**. In this introduction, I will provide the necessary background knowledge and the overarching questions that drove my dissertation research. Additional background information, more specific to each project, is also provided in the introduction section of each subsequent chapter.

1.2 Human innate immune response to infection

1.2.1 Innate immune response during acute infection

Humans encounter millions of potentially pathogenic microorganisms daily from the food we eat, objects we touch and the air we breathe. To quickly identify and eliminate these overwhelming invading pathogens, we rely on our innate immune response. In almost all somatic cells, the initiation of the innate immune response is dependent on the recognition of pathogen-associated molecules that mark "self" from "non-self", such as the lipopolysaccharide (LPS) on gram-negative bacteria and unique double-stranded RNA from certain viruses

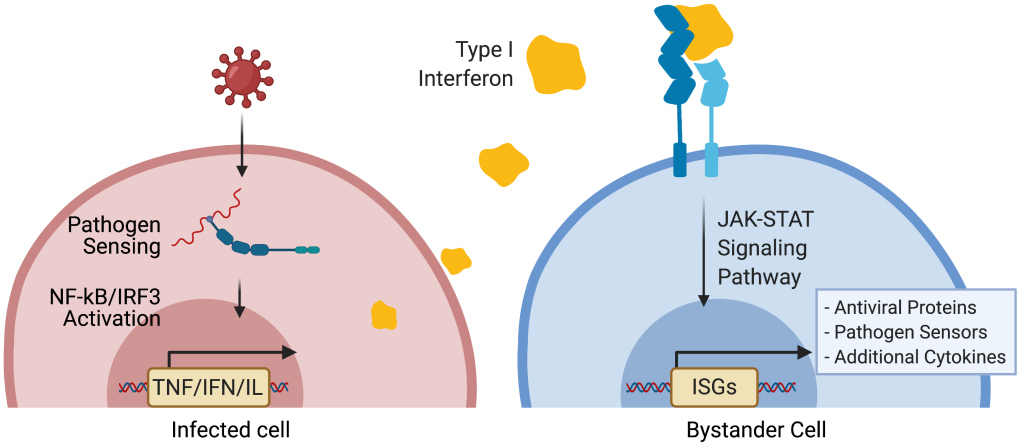


Figure 1.1: **Diagram summarizing pathogen sensing and interferon responses of the innate immunity.** Upon infection, pathogen-specific molecular signatures are detected by extracellular/intracellular pattern recognition receptors. Upon activation by pathogen signatures, these receptors initiate the signaling cascade, leading to the activation of inflammatory transcription factors, such as nuclear factor kappa B (NF-κB) and interferon regulatory factor 3 (IRF3). This triggers the expression of inflammatory genes along with the release of intercellular signaling molecules, including interleukins (IL), tumor necrosis factor (TNF) and interferons (IFN). In the case of bystander cells, upon the binding of IFN, the janus kinase (JAK)- signal transducer and activator of transcription proteins (STAT) pathway is activated. This in turn results in the transcription of interferon stimulated genes (ISGs) that are responsible for downstream anti-pathogen responses.

[5]. The recognition of these "non-self" molecules subsequently activates the inflammatory pathway, which result in the release of key inter-cellular signaling molecules including tumor necrosis factors (TNF), interleukins (IL), and interferons (IFN) (Figure 1.1). These signaling molecules are the key to transforming surrounding cells into the anti-pathogen state, and to recruiting immune cells to the site of infection, both being equally important for containing the ongoing infection.

Specifically, Type-I interferon secreted by the infected cell is essential to prepare both the infected and surrounding bystander cells for an upcoming infection. Upon the binding of the type-I interferon to its receptor, janus kinases (JAK) and signal transducer and activator of transcription proteins (STAT) orchestrate a series of transcription activation events that lead to the increased expression of a group of genes that are known to directly detect and restrict invading pathogens, and modulate the cellular homeostasis to create a hostile environment for the pathogen replication [6] (Figure 1.1). The elements of the interferon-stimulated response are molecular keys to pathogen restriction. However, due to cell-type specificity and the differences in response triggered by different pathogens, it remains unknown if there is a core response that is consistent across all cell types and different infections [7]. To address this question, I compared transcriptional response elicited by diverse cell types undergoing different infections (see **Chapter 4**). I uncovered a group of genes whose expressions are tightly correlated with the infection status, independent of cell and pathogen types. This group of universal response genes could help us understand the immune transcriptional regulation, as well as be used as signatures for infection diagnostics, as further explained in the Introduction sections below.

1.2.2 Genome-wide profiling of innate immune response

Transcriptomic profiling of innate immune response by RNA-seq

One of the most fundamental questions for understanding the innate immune response is: "What genes are expressed during the infection?". In the past decade, high throughput

Illumina sequencing has become widely adapted to answer such question [8]. Specifically, RNA sequencing (RNA-seq), which involves reverse transcribing messenger RNA from a biospecimen into complementary DNA followed by sequencing, has been used extensively to probe gene expression [9]. In RNA-seq, each read is a sampling from a single mRNA. These reads, which often number in the millions, are mapped back onto a reference sequence (Figure 1.2). As a highly expressed gene would generate more mRNA, and hence more sequencing reads would originate from that mRNA, one can obtain a quantitative readout of the gene expression across the entire genome.

Adapting the RNA-seq technique to studies of human innate immune response to infection, researchers compare the gene expression between an infected and mock-infected specimen (infected by the pathogen-of-interest), or between specimen treated/untreated with immune signaling molecules. The comparison then yields statistically significant differences in gene expression between the two conditions, a technique termed differential expression analysis. In brief, the sequencing data from each experiment is first normalized to its relative library size. Then the dispersion of the expression across the replicates are estimated by accounting for all the experimental replicates. Finally, a statistical test (such as Wald test implemented in the R package DESeq2 [10]) estimates both the relative fold change between the two treatment conditions and statistical significance (typically p-values) from the hypothesis testing. Using this technique, I first conducted a meta-analysis of assorted RNA-seq datasets obtained from publicly available database to look for genes that are universally up-regulated upon diverse infections within different cell types. I hence collected human saliva samples from infected/healthy individuals, generated RNA-seq datasets myself, and confirmed that the majority of these universal response genes are also over-expressed in infected saliva samples (see **Chapter 4**). I also generated RNA-seq datasets that profiled both blood- and induced pluripotent stem cell (iPSC) derived-macrophages during their differentiation and polarization process (additional description of this work is described in the following Introduction sections and in **Chapter 5**).

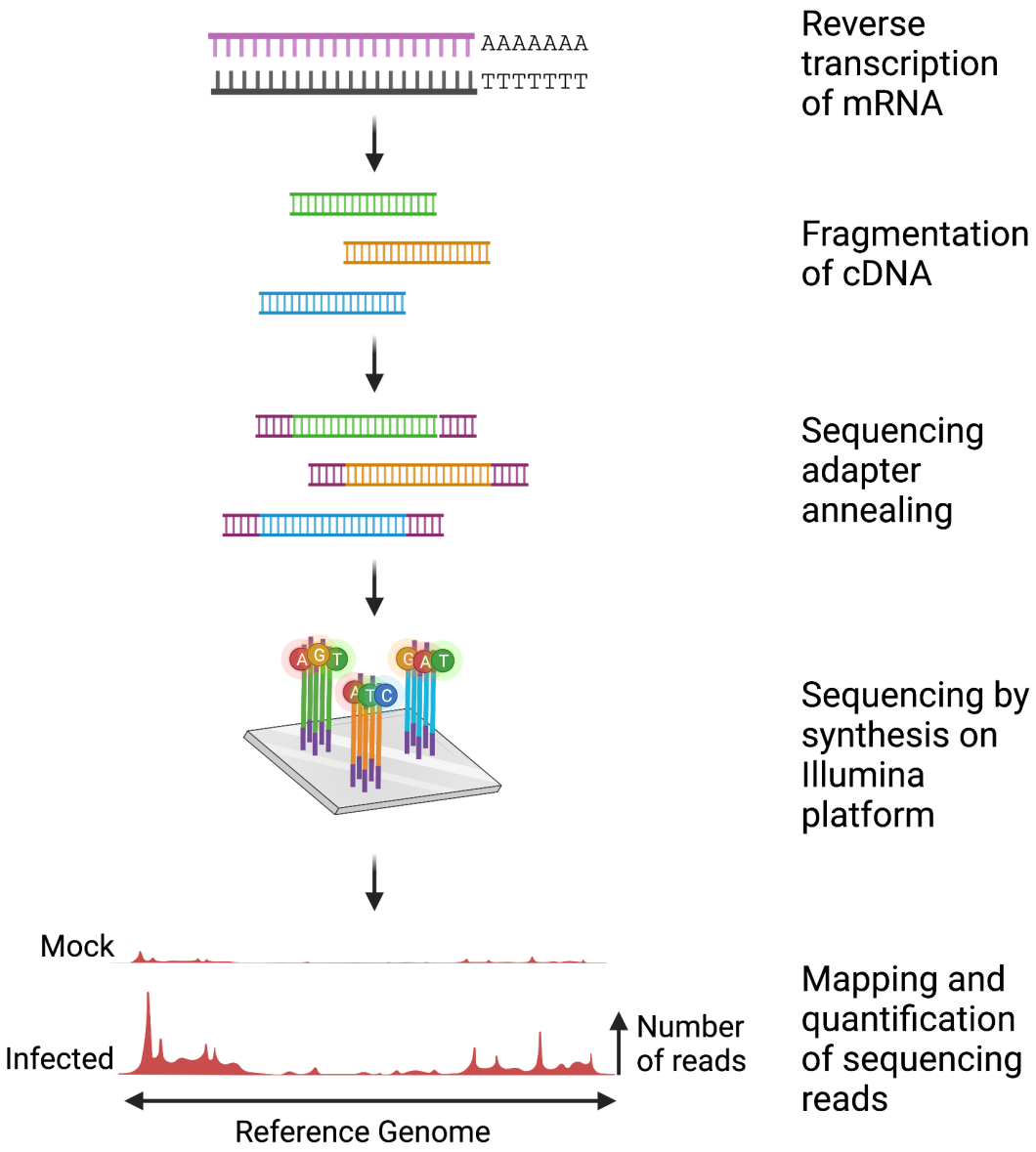


Figure 1.2: **Quantification of gene expression through high-throughput RNA sequencing.** To quantify the genome-wide gene expression, messenger RNA (mRNA) is harvested from cells or biospecimen and reverse transcribed into complementary DNA (cDNA). The cDNA is then sheared into shorter fragments, and sequencing adapters are ligated to the fragmented cDNA (based on the specific library construction kit, these steps might be in a different order). The cDNA library is then sequenced by synthesis on Illumina platforms, resulting in millions of short sequencing reads. These sequencing reads are then quality-filtered and mapped onto the reference genome with proper gene annotations. Using gene annotations, the amount of reads mapped onto specific genes is tallied and reflects the relative gene expression levels, which can be compared across different treatment conditions to infer differential gene expression.

Inferring transcriptional programming of innate immune response via ATAC-seq

Beyond identifying the gene expression changes observed during infection or the presence of immune signaling molecule, it is important to understand how these genes are tightly regulated. A recently-developed technique, Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq) enables the examination of the open chromatin regions across the genome, and hence infer the underlying transcriptional regulation [11](Figure 1.3). Briefly, ATAC-seq is carried out by using a hyperactive transposase to detect and liberate open chromatin regions in the genome. These liberated genomic DNA are then processed for high-throughput sequencing. The resulting sequencing reads are then mapped back to the reference genome to reveal which regions of the genome reside in open chromatins, as more mapped sequencing reads are often associated with a more accessible genomic region. Based on the observation that some transcription factors binding changes the chromatin accessibility, one can then infer the activity of many transcription factors based on the ATAC-seq read abundance as well as the binding motifs embedded in the genomic region [12, 13]. Using this approach, I profiled the chromatin landscape to identify changes observed in both blood- and iPSC-derived monocytes/macrophages under treatment of different immune signaling molecules (see **Chapter 5**). From these datasets, I then inferred which unique transcription factors are essential for monocyte/macrophage differentiation and polarization.

1.3 The role of macrophages during infection

1.3.1 Immunological functions of macrophages

As mentioned in the section above, beyond the autocrine and paracrine interferon signaling, one crucial aspect of the innate immunity is the recruitment of immune cells. Macrophages, along with neutrophils, NK cells and dendritic cells, are the key players that form the next line of defense against the invading pathogens [14–16]. Specifically, macrophages play an essential role in modulating this defense by directly phagocytizing the

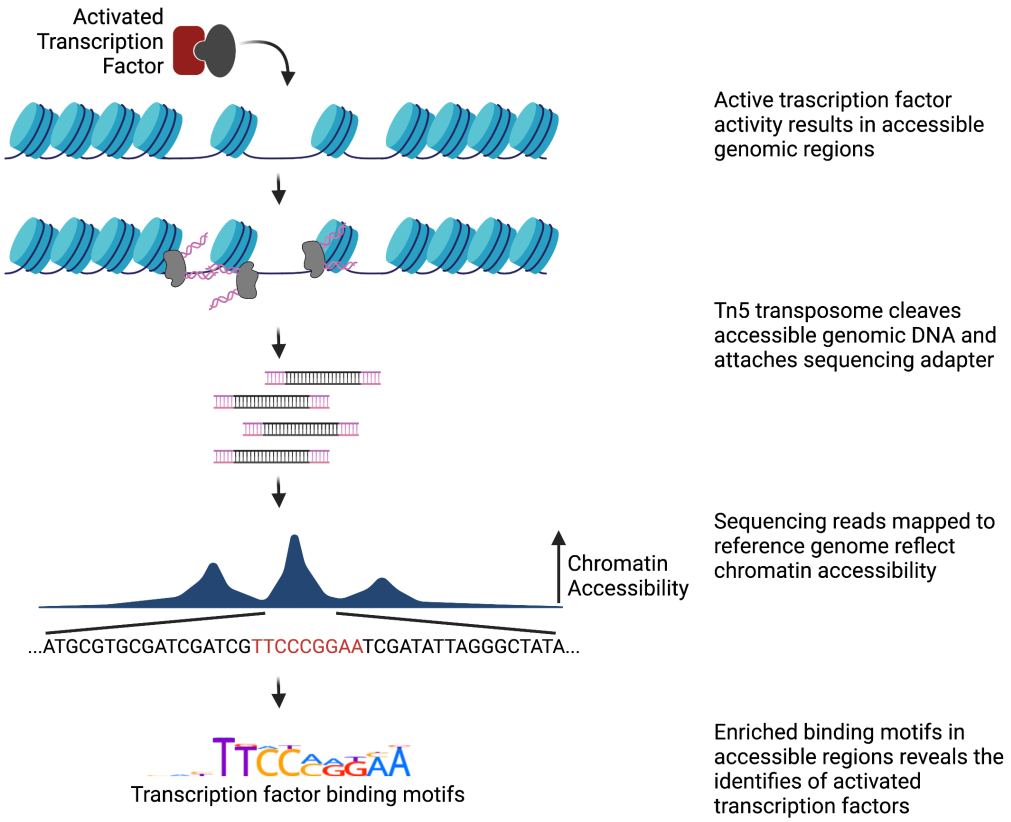


Figure 1.3: **Inferring transcription regulation via ATAC-seq.** To understand which transcription factors are active within specific cell type or during innate immune response, the assay for transposase-accessible chromatin sequencing (ATAC-seq) can be used. This is based on the assumption that genome regions become more accessible upon active transcription bindings [13]. The accessible regions of the genome are then cleaved and ligated with sequencing adapters in one step using the Tn5 transposome. These DNA fragments are then sequenced through the Illumina platform, and the resulting sequencing reads can be used to identify and quantify chromatin openness genome wide. The DNA sequences from the accessible genomic regions can then be used to identify transcription factor binding motifs. Combining the identity of the transcription factor along with the chromatin accessibility, one can infer the relative transcription factor binding activities.

pathogen or infected cells, and by presenting the digested pathogen's molecular features to cytotoxic T cells and T helper cells to initiate the downstream adaptive immune response.

Throughout our body, some macrophages are tissue-resident, mediating homeostasis and providing immune surveillance at the specific tissue, while other macrophages are derived from blood circulating monocytes upon immune activation. Previous studies have pin-pointed the key signaling molecule, macrophage colony forming factor (M-CSF), that triggers the monocyte-macrophage differentiation [16]. Upon differentiation and arrival at the site of infection, these macrophages can hence be polarized towards inflammatory (M1) or anti-inflammatory (M2) subtypes given the presence of additional molecular cues [17] (Figure 1.4). This enables the macrophages to adapt to the disease progression, mediating pathogen defense (M1) or tissue repair (M2). Experimentally, researchers have also devised *in vitro* conditions that triggers the macrophage polarization, using interferon-γ and LPS for M1 polarization, and IL-4 and/or IL-13 for M2 polarization [16]. However, the gene expression changes during the macrophage polarization has not been systematically examined. Using the RNA-seq and ATAC-seq techniques mentioned in the previous section, I explored such changes at both transcriptional and regulatory levels in both primary and iPSC-derived macrophages (see **Chapter 5**).

1.3.2 Macrophages as viral targets

Upon activation, blood-circulating monocytes become macrophages and are migratory towards the site of infection, these characteristics are often hijacked by pathogens for systemic dissemination, persistent infection and replication [18, 19]. In our lab, we study several viruses of great clinical significance, including HIV-1, dengue virus (DENV) and influenza virus. All three viruses have been previous reported to infect monocytes/macrophages, leading to severe infection outcomes [20]. Specifically, macrophages support productive HIV-1 infection, which in turn trigger the secretion of proinflammatory cytokines that attracts uninfected CD4+ T cells, accelerating the virus dissemination [21]. In fact, HIV-1 infected

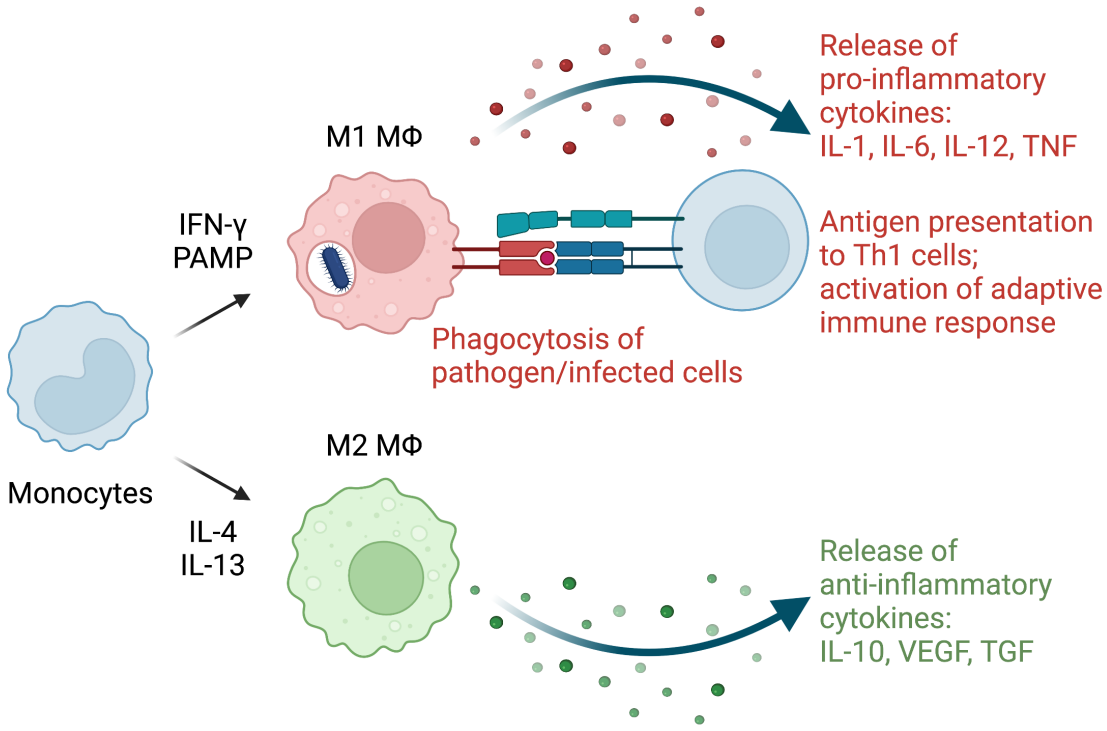


Figure 1.4: **Overview of macrophage functions during immune response.** Upon immune activation, blood-circulating monocytes can arrive at the site of infection and differentiate into M1 macrophages in the presence of pathogen-associated molecular patterns (PAMP) and interferon (IFN) secreted by the infected cells. The M1 macrophages are responsible for mediating the immune response through direct phagocytosis, release of additional pro-inflammatory cytokines that recruit additional lymphocytes, as well as antigen presentation to activate T cells, which eventually leads to the activation of adaptive immune response. On the other hand, upon the clearing of infection, M2 macrophages then mediate tissue repair through the release of anti-inflammatory cytokines. Additional abbreviation used in the figure: IL: interleukin, MΦ: macrophage, TNF: tumor necrosis factor, VEGF: vascular endothelial growth factor, TGF: transforming growth factor beta.

macrophages are found in spleen, lung, heart, colon, and brain, suggesting their critical role as a HIV-1 reservoir during persistent infections [22]. In the case of dengue virus infection, macrophages are not only hijacked by the viruses to establish systemic infection, but the virus also utilizes the macrophage Fc receptors to achieve antibody-dependent enhancement during the secondary infection This often leads to severe dengue fever in infected hosts [23–25]. Finally, highly pathogenic influenza viruses are also shown to productively replicate in macrophages [26–28]. Such interaction is shown to trigger high-level production of proinflammatory cytokines by the infected macrophages, and is thought to be the cause of severe pneumonia in infected hosts [29]. Together, this highlights the importance of using macrophages as a model system to study these pathogen-host interactions. During my dissertation research, I demonstrated that all three viruses are fully capable of infecting both primary and iPSC-derived macrophages *in vitro*. Using flow cytometric techniques, I further demonstrated how macrophages have potential in restricting lowly pathogenic (seasonal) influenza virus infection at late stage of virus life cycle (see **Chapter 5**).

1.3.3 Methods for deriving macrophages

Traditionally, in order to study pathogen-human macrophage interaction *in vitro*, researchers often have to rely on blood donation, from which they purify blood-circulating monocytes via sequential gradient centrifugations and subsequently differentiate them into macrophages via M-CSF treatment [30] (Figure 1.5). This process requires a complicated institutional review board approval, and heavily relies on the availability of the blood donors. Even with everything in place, the quantity of blood-derived monocytes is very limited, as they only constitute ∼10% of circulating immune cells, i.e. only 2-5 millions of monocytes can be isolated from 50 mL of blood draw [31]. Blood-derived monocytes also suffer from inter-individual variability (differences between donors) as well as the lack of genome-editing tools, as monocytes/macrophages are not proliferative.

One possible alternative to blood-derived monocytes is immortalized monocytic cell

lines. These cell lines, such as THP-1, U-937 and HL-60, are derived from patients with monocytic leukemia or histiocytic lymphoma [32] (Figure 1.5). Compared to the primary cells, these immortalized cell lines are proliferative, so that they can be expanded to large numbers for complex experiments. However, numerous studies have reported the lack of relevancy of these cell lines: they do not respond to cytokine activation to the similar degree, nor do they posse a similar level of restriction against viral infections, as compared to their blood-derived counterpart [27, 33–35]. Together, this drove me to seek for an alternative monocyte/macrophage model.

Induced pluripotent stem cells (iPSCs) are somatic cells reprogrammed to exhibit pluripotency via forced expressions of several transcription factors (Oct4, Sox2, Lin28, Klf4, and L-Myc) [36, 37]. Through the treatment of specific combinations of signaling molecules, iPSCs have been successfully differentiated into neurons, myocytes and lymphocytes as relevant model systems [38]. Recently, iPSC differentiation into monocytes was also made possible via a serial feeder-free system: first, bone morphogenetic protein 4 (BMP4), vascular endothelial growth factor A (VEGFA), and GSK-3 inhibitor were supplemented to iPSC for mesoderm differentiation; next, the mesoderm is further differentiated into hematopoietic progenitors with the addition of transforming growth factor (TGF)-beta inhibitor, stem cell factor (SCF), thyroid peroxidase (Tpo), IL-3, and FLT-3 Ligand (FL); finally, the monocytes were derived by adding M-CSF and granulocyte-macrophage colony-stimulating factor (GM-CSF) [39, 40](Figure 1.5). Using a similar paradigm, we successfully differentiated monocytes from two iPSC cell lines. This approach allowed us to harvest large quantities (40-50 millions) of monocytes within one round of differentiation. I then carried out phenotypical, transcriptional and functional comparisons between the blood- and iPSC-derived monocytes/macrophages and found they are virtually identical at all levels of comparison (see **Chapter 5**). Since iPSCs are also genetically modifiable, this provides a valuable alternative model system for studying monocyte/macrophage-pathogen interaction.

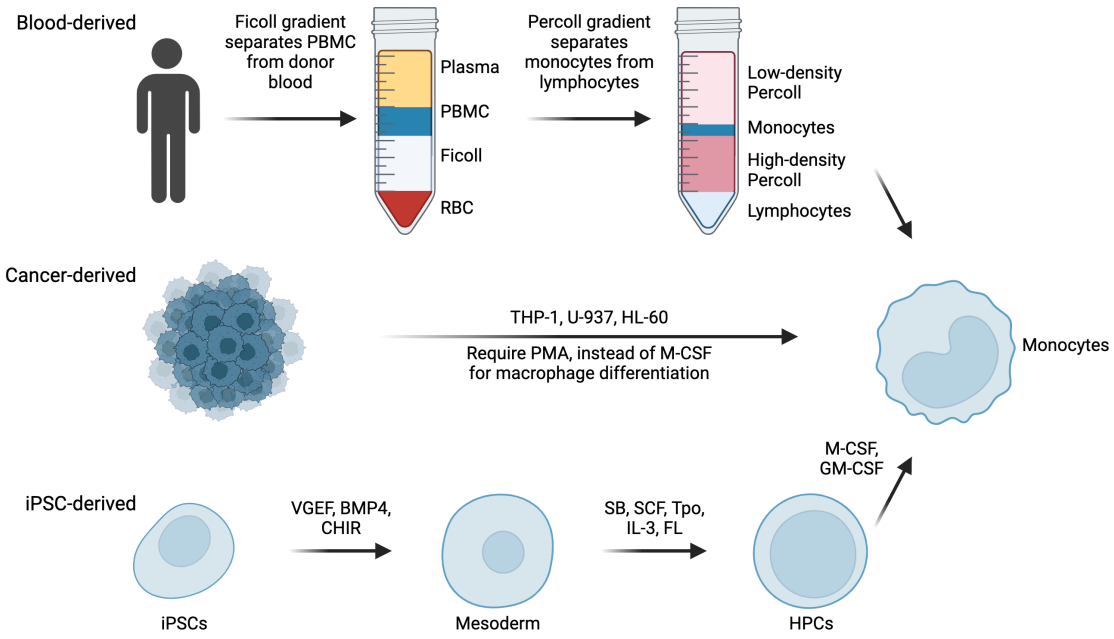


Figure 1.5: **Overview of methods for deriving monocytes through different sources.** Blood-derived monocytes are directly obtained through blood donation, from which the peripheral blood mononuclear cell (PBMCs) are separated from plasma and red blood cells (RBCs) using the Ficoll gradient. Subsequently, the PBMCs are loaded onto gradient Percoll column, which separates monocytes from lymphocytes based on their size differences [30]. Immortalized cancer-derived monocytic cells are often used as an alternative to the primary monocytes, but differentiation into macrophage often requires synthetic drug tetradecanoylphorbol acetate (PMA) instead of more biologically relevant macrophage colony forming factor (M-CSF). iPSC-derived monocyte is a newly developed model system that more closely resembles primary monocytes. The differentiation from iPSC requires several stages through mesoderm, hematopoietic stem cell (HSCs) then to monocytes using a combination of different growth factor and cytokine cocktails. Additional abbreviations in the figure: BMP4: bone morphogenetic protein 4, VEGF: vascular endothelial growth factor, CHIR: GSK-3 inhibitor, SB: TGF-beta/ALK inhibitor, SCF: stem cell factor, Tpo: thyroid peroxidase, FL: FLT-3 Ligand, GM-CSF: granulocyte-macrophage colony-stimulating factor.

1.4 Diagnostics for infection

1.4.1 Types of infection diagnostics

Another essential aspect of pandemic response is the surveillance of infected individuals. In order to monitor and contain disease outbreaks, effective infection diagnostics become the most fundamental tool in the public health arsenal. Currently, infection diagnostics can be categorized based on the type of molecules being detected: (1) pathogen genome-derived nucleic acid, (2) pathogen-derived protein, and (3) pathogen-specific antibody [41, 42].

As these molecular signatures tend to arise at different stages of infections, proper usage of the correct type of diagnostics at the correct time dictates the detection sensitivity. Both pathogen-derived nucleic acid and protein (antigen) assays require the actively replicating pathogens to reach a certain abundance [42]. As nucleic acid-based detection usually involves amplification, it typically outperforms the antigen-based assays near the detection limit [43]. In contrast, antigen-based assays excel in rapid turnaround time and at-home usability. Moreover, as different pathogens have distinct primary sites of infection (due to the host receptor-specificity, interactions with host immune system and routes of transmission), the choice of diagnosis biospecimen hence needs to adapt to our understanding on the pathogen's pathogenesis. On the other hand, diagnostics that rely on host-derived antibodies against particular pathogens are primarily used to examine historical infections [44]. While such information is beneficial in helping clinicians determine the cause of symptoms, it is ineffective in controlling the spread of infectious diseases because the transmission likely has already happened. During the COVID-19 pandemic, all three forms of diagnostics were widely used. We learned important lessons that the types of diagnostics, the testing deployment strategy, along with the adaptive public policies are all crucial to control the transmission. However, several important questions remain: (1) How can the nucleic-acid based assay be faster and more adaptable to at-home use; (2) How can we detect infected individuals at an earlier time point of infection, ideally when one is asymptomatic or pre-

symptomatic; (3) How can we detect diverse types of infections with a single assay? My dissertation first highlighted the prevalence of asymptomatic SARS-CoV-2 infection within the local population see **Chapter 2**, and then sought to address these problems via a saliva-based SARS-CoV-2 screening test and an early agnostic infection diagnostic paradigm in **Chapter 3** and **Chapter 4**, respectively.

1.4.2 Values of saliva in infection diagnostics

During the COVID-19 pandemic, saliva emerges as an alternative biospecimen with high diagnostic values, and it was used for SARS-CoV-2 screening in a large population [45–48]. Comparing to blood and nasopharyngeal/throat swabs, saliva collection is less invasive and can be carried out independently by the donor, minimizing healthcare worker infections during sample collection. It has been shown that most of the respiratory pathogen genomes (SARS-CoV-2, SARS-CoV, influenza virus, respiratory syncytial virus, rhinovirus, and *M. tuberculosis* etc.) are readily present in saliva samples of infected individuals [46, 49–54]. As these pathogens replicate in the respiratory tract, either through direct invasion into oral tissues or through indirect infection in nasal/lung epithelial cells, the intact pathogen or pathogen fragments are often released into the respiratory fluid, leading to airborne transmission and/or detection of pathogen genomes in the saliva. This led to the development of multiple saliva-based multi-pathogen diagnostic assays, which allowed clinicians to narrow down the cause of illness with just a small amount of patient saliva [55–57]. Moreover, it has even been shown that the genomes of some blood-borne pathogens (dengue virus, HIV, and Ebola virus) along with host-derived antibodies can also be detected in saliva, leading to follow-up studies that indicated the origin of saliva is partially blood-derived plasma filtered through the salivary glands. These findings further highlight the additional diagnostic value of saliva [58–62].

Beyond detecting pathogen genomes, saliva has been often described as the mirror of the body, as it harbors diverse extracellular vesicles, endocrine signaling molecules, and

secretory antibodies [63]. Thus saliva has the potential for monitoring systemic homeostasis. In fact, saliva has already been widely used to detect the presence of oral squamous cell carcinoma, the main type of oral cancer [64, 65]. In oral cancer patients, it has been shown that inflammatory cytokines, such as tumor necrosis factor α, IL-1 and IL-6, along with the cytokine-induced transcriptional changes, are readily detectable in saliva as the result of innate immune response against the oral cancer cells. In addition, saliva is also used for the diagnosis of Sjogren syndrome, an autoimmune disease that is characterized by the chronic infiltration of lymphocytes into salivary gland [66]. Autoantibodies and inflammatory cytokines present in saliva are reliable evidence for the Sjogren syndrome. Remarkably, in studies characterizing the saliva and blood-level immune response during these oral diseases, high concordance was found between the local (saliva) and systemic (blood) response [64–66]. However, the application of using saliva for diagnosis beyond the scope of oral diseases is very limited. Given the systemic response is readily captured in saliva, it is reasonable to believe that saliva holds much greater diagnostic values for diseases incurred throughout our body.

1.4.3 Using host immune signatures of infection diagnostics

As mentioned in the previous sections, the COVID-19 pandemic also highlights the need for new diagnostic approaches that are early and agnostic: A diagnostics that identifies infected individuals before the onset of symptoms could drastically reduce the risk of subsequent transmission, while a diagnostic independent of infection type would ensure logistic readiness during the outbreak of a novel pathogen. Many attempts have been made to develop agnostic diagnostics by incorporating a panel of pathogen targets into the assay [55–57]. However, the ability to detect any pathogen present in the biospecimen still relies heavily on the abundance of the pathogen signatures, which, in most of the cases, tracks with symptoms and the increased pathogen transmissibility at the same time.

Human innate immune response has evolved to counteract an ongoing infection with

speed and versatility: it is activated at the time of pathogen exposure, and the activating signal is propagated systemically through blood-circulating cytokines [5]. Upon activation, the innate immune response ubiquitously neutralizes the pathogen or infected cells through recruitment of phagocytic cells and programmed cell death. The systemic cytokine release also triggers rapid expression changes in nearly all affected somatic cells result in the production of anti-pathogen proteins as well as additional cytokine for signal amplification [6]. Understanding that: 1) innate immune response is activated immediately upon exposure, 2) the activation of innate immune response is universal during almost all infections, and 3) the signal of innate immune response is amplified through systemic cytokine signaling, I reason that the characteristics of human innate immune response are ideal signatures for the next-generation early-and-agnostic diagnostics.

Previous studies have monitored the transcription changes observed in innate immune response in blood, and found that these changes correlate well with infection status, disease severity, and symptomatic status [67–72]. Based on these findings, blood-based immune activation monitoring assays have been developed for clinical uses, especially for the diagnosis of sepsis [73, 74]. However, blood as an invasive biospecimen carries a behavior barrier so that blood-based diagnostics are predominantly carried out in a clinical setting. Based on the noninvasiveness nature of saliva, I explored the possibility of detecting these innate immune response signatures within saliva as a potential alternative for future infection screenings (see **Chapter 4**).

In summary, in this introduction chapter, I provided background information that is fundamental to the rest of my dissertation. What follows are chapters that summarize my past projects: In **Chapter 2**, I highlighted the prevalence of asymptomatic SARS-CoV-2 carriers, and the importance of infection screening in the seemingly healthy population during the pandemic. To address such challenge, I first developed a saliva-based SARS-CoV-2 screening assay with fast turnaround time (**Chapter 3**). Subsequently, to better prepare for the future disease outbreak, I identified the universal transcriptional response

from human innate immunity as signatures for infection, and showed that such signatures are readily detectable in saliva (**Chapter 4**). Finally, to enable more relevant future studies of host response to infection, I validated an iPSC-derived macrophage model system via comprehensive phenotypical, transcriptomic and functional analyses (**Chapter 5**).

Chapter 2

Just 2% of SARS-CoV-2-positive individuals carry 90% of the virus circulating in communities

---

Adapted from: Yang Q, Saldi TK, Gonzales PK, Lasda E, Decker CJ, Tat KL, Fink MR, Hager CR, Davis JC, Ozeroff CD, Muhlrad D, Clark SK, Fattor WT, Meyerson NR, Paige CL, Gilchrist AR, Barbachano-Guerrero A, Worden-Sapper ER, Wu SS, Brisson GR, McQueen MB, Dowell RD, Leinwand L, Parker R, Sawyer SL. Just 2% of SARS-CoV-2-positive individuals carry 90% of the virus circulating in communities. Proceedings of the National Academy of Sciences. 2021 May 25;118(21):e2104547118.

2.1 Contributions

This chapter describes the collective work of our University COVID-19 response team that worked tirelessly in screening the entire university population for SARS-CoV-2 during the pandemic. I was initially involved in this response, and tasked with setting up the screening lab and designing the layout of the sample collection sites. Later during the semester, I helped organizing the collected samples and the SARS-CoV-2 screening RT-qPCR data. Using this valuable collection of screening data, I explored the campus population viral load distribution and summarized the results in this manuscript. However, I was by no means alone in this process. This response is the collective work of everyone listed here as the co-authors. I would especially like to dedicate this manuscript to our beloved colleague, Denise Muhlrad, who unfortunately passed away while performing her essential duty in the lab.

2.2 Abstract

We analyze data from the Fall 2020 pandemic response efforts at the University of Colorado Boulder (USA), where more than 72,500 saliva samples were tested for SARS-CoV-2 using quantitative RT-PCR. All samples were collected from individuals who reported no symptoms associated with COVID-19 on the day of collection. From these, 1,405 positive cases were identified. The distribution of viral loads within these asymptomatic individuals was indistinguishable from what has been previously observed in symptomatic individuals. Regardless of symptomatic status, approximately 50% of individuals who test positive for SARS-CoV-2 seem to be in non-infectious phases of the disease, based on having low viral loads in a range from which live virus has rarely been isolated. We find that, at any given time, just 2% of individuals carry 90% of the virions circulating within communities, serving as viral "super-carriers" and possibly also super-spreaders.

2.3 Introduction

SARS-CoV-2 is a novel coronavirus that emerged into the human population in late 2019 [75], presumably from animal reservoirs [76, 77]. During the ensuing world-wide pandemic, already more than two million lives have been lost due to the virus. Spread of SARS-CoV-2 has thus far been extremely difficult to contain. One key reason for this is that both pre-symptomatic and asymptomatic infected individuals can transmit the virus to others [76, 78–86]. Further, it is becoming clear that certain individuals play a key role in seeding super-spreading events [87–90]. Here, we analyzed data from a large university surveillance program. Viral loads were measured in saliva, which has proven to be an accessible and reliable biospecimen in which to identify carriers of this respiratory pathogen, and the most likely medium for SARS-CoV-2 transmission [45, 91, 92]. Our dataset is unique in that all SARS-CoV-2-positive individuals reported no symptoms at the time of saliva collection, and therefore were infected but asymptomatic or pre-symptomatic. We find that

the distribution of SARS-CoV-2 viral loads on our campus is indistinguishable from what has previously been observed in symptomatic and hospitalized individuals. Strikingly, these datasets demonstrate widespread differences in viral levels between individuals, with a very small minority of the infected individuals harboring the vast majority of the infectious virions.

2.4 Results

2.4.1 The University of Colorado Boulder SARS-CoV-2 screening operation

We analyzed data resulting from SARS-CoV-2 testing performed on the University of Colorado Boulder (USA) campus during the Fall academic semester of 2020 (Aug. 27th – Dec. 11th, 2020). Residents of dormitories were tested weekly, and several campus testing sites were in operation throughout the semester offering testing for any campus affiliate. At the time of saliva collection participants were asked to confirm that symptoms were not present; therefore, any infected persons identified through this surveillance testing were asymptomatic or pre-symptomatic at the time of saliva collection. It should be noted that all of the samples analyzed herein were collected before the B.1.1.7 ("U.K.") SARS-CoV-2 variant, and subsequent major variants of concern, were first documented in the United States during the final weeks of 2020 and the beginning of 2021 [93].

During the Fall 2020 semester, more than 72,500 saliva samples were screened for SARS-CoV-2. A quantitative reverse transcription-polymerase chain reaction (RT-qPCR) assay was used, with the template coming from the direct addition of saliva without RNA purification [94]. Three TaqMan primer/probe sets were used in a multiplex reaction directed against two regions of the SARS-CoV-2 genome (CU-E and CU-N, where CU stands for the University of Colorado) and a host transcript (CU-RNaseP) as control. The multiplex reaction was used to create standard curves to convert Ct value (cycle threshold) of each primer set to viral load (virions per mL) in the original saliva sample (Fig. 2.1A). To

ensure the viral load quantification is accurate for samples with extremely low Ct values (i.e. extremely high viral loads), we performed serial dilution of three saliva samples with amongst the highest observed viral loads of the semester, and showed that Ct values scale linearly with the dilution factor (Fig. 2.1B).

From over 72,500 saliva samples screened, 1,405 SARS-CoV-2-positive samples were identified. The vast majority of these positive samples were from unique individuals, because individuals with positive tests were directed into the healthcare system for further testing and care. The distribution of the Ct values of these 1,405 individuals, with each of the two primer sets used, is shown in Fig. 2.2A. Overall, the distribution of SARS-CoV-2 viral load fits under a log-normal distribution centered around the mean of $9.9 * 10^8$ virions/mL (median = $1.1 * 10^6$ virions/mL) for the CU-E primers or $3.6 * 10^8$ virions/mL (median = $2.5 * 10^5$ virions/mL) for the CU-N primers (Fig. 2.6). The highest observed viral load was over 6 trillion ($6.1 * 10^{12}$) virions/mL, which was only observed in one individual. It is remarkable to consider that this individual was on campus and reported no symptoms at our testing site. The lowest viral load detected was 8 virions/mL. Thus, surveillance testing demonstrates an extremely wide variation in the viral load in infected but seemingly healthy (asymptomatic) individuals.

To verify that these viral load distributions were not influenced by the specific RT-PCR primers used, we determined the agreement between the CU-N and CU-E primers with regard to the Ct values produced from samples. Different primer sets should be expected to produce slightly different Ct values on the same samples, due to differences in primer efficiencies and human pipetting error during reaction setup. Nonetheless, we find a tight correlation in samples with Ct values < 30 (Pearson correlation coefficient between CU-N and CU-E Ct values = 0.92), but this correlation breaks down in samples with higher Ct values (Pearson correlation coefficient between CU-N and CU-E Ct values = 0.10, Fig. 2.2B). At high Ct values (i.e. low viral loads) weaker correlation is likely a result of stochasticity in reverse transcription and/or in the initial rounds of PCR. This is supported by an in-

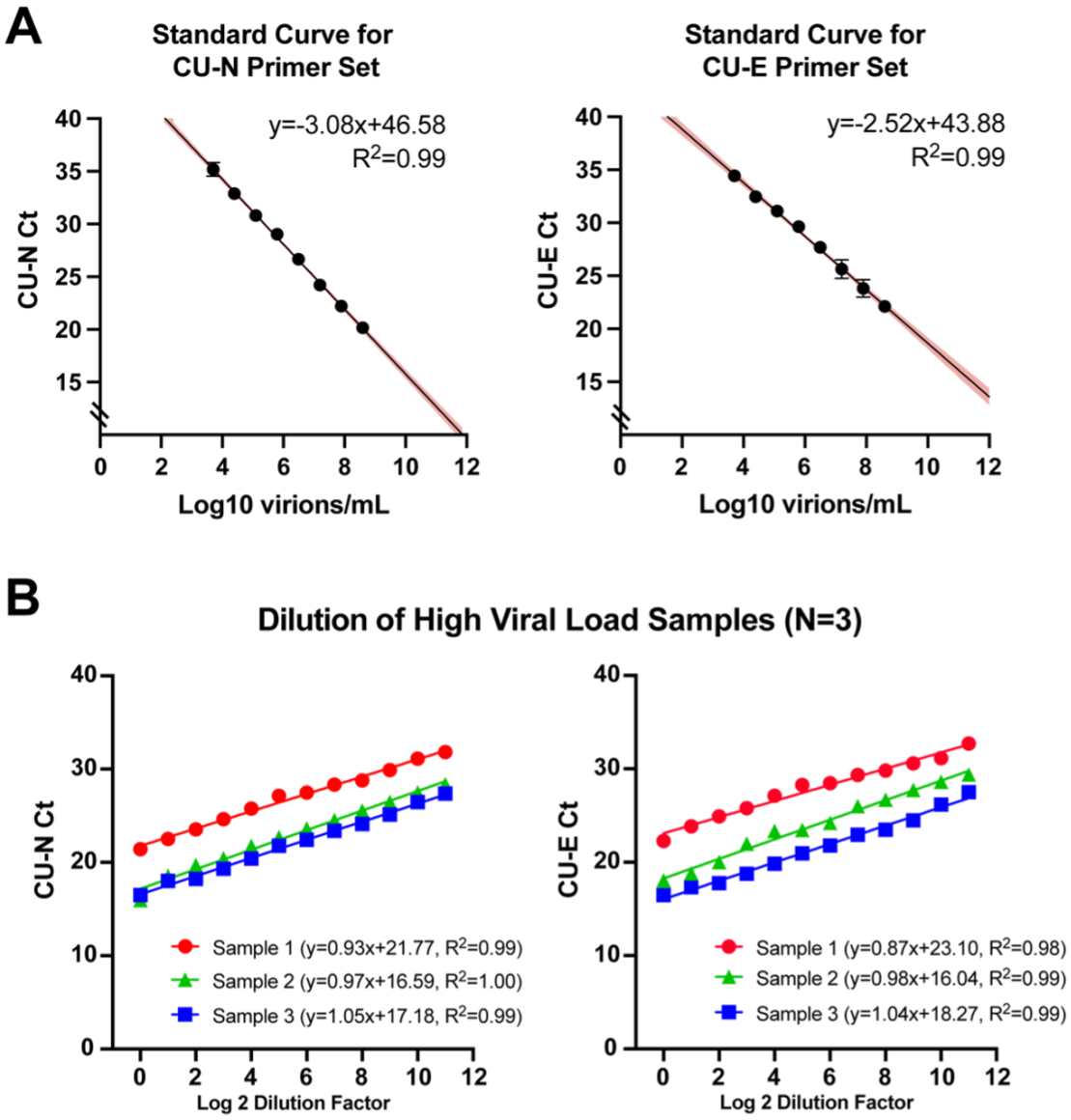


Figure 2.1: **Examination of the linearity of the RT-qPCR assay over the range of viral loads observed in this study (A)** Standard curves were created for the primer sets used in this study. Gamma-irradiated SARS-CoV-2 virions (BEI Resources NR-52287) were spiked into healthy saliva samples from three different individuals to reach $4*10^8$ virions/mL and incubated for 30 minutes at 95^{o}C. Samples were then serially diluted 1:5 using heat-treated healthy saliva from the same individuals as the diluent, yielding the indicated final concentrations (X-axis). Samples were then subjected to the multiplex RT-qPCR reaction described in the method. The standard curve for each primer set was generated by linear regression analysis of the triplicate experiment and is illustrated with 95% confidence interval (R-squared $>$ 0.99 for both standard curves). **(B)** We performed serial dilution of three of the saliva samples with amongst the highest observed viral loads of the semester ($8.1 * 10^8$ – $1.2 * 10^{11}$ virions/mL) to determine the linear range of the RT-qPCR assay. Saliva was incubated for 30 minutes at 95^{o}C, then diluted 1:2 in series using heat-treated healthy saliva as the diluent. Linear regression was performed on the dilution series to show that the Ct values scale linearly with dilution factors (R-squared $>$ 0.98 for all dilution curves).

depth analysis performed on 105 of the SARS-CoV-2 positive samples, where each sample was analyzed with eight different primer sets commonly used in SARS-CoV-2 diagnostic tests (Fig. 2.2C and Fig. 2.5). We see tight congruence between Ct values generated with different primers on the same samples, especially at Ct values < 30. Overall, since the CU-E primer set demonstrated the highest consistency with other primer sets during this in-depth comparison (Fig. 2.5), we used the Ct values resulting from this primer set to calculate saliva viral loads from this point forward.

2.4.2 Populations have similar viral load distributions regardless of symptomatic status

We next compared viral loads from individuals on our campus, who had no symptoms at the time of sample collection, to similar viral load measurements taken in saliva of symptomatic individuals. We examined published SARS-CoV-2 RT-qPCR datasets derived from studies of hospitalized (and therefore symptomatic) individuals. We specifically sought studies that assayed saliva and where viral loads were reported, since Ct values are lab- and assay- specific (23). We identified 404 datapoints that met such criteria, which we collated from the 10 studies listed in Supplementary Table 2.1. We note that our campus sampling likely represents earlier average timepoints in the course of infection than that of the hospital samples, which were mostly collected after symptom onset. Nonetheless, similar to the viral load distribution of the campus asymptomatic population (mean $= 9.9 * 10^8$ virions/mL, median $= 1.1 * 10^6$ virions/mL), the viral load in symptomatic patient saliva samples shows a log-normal distribution with a mean of $1.8 * 10^9$ virions/mL (median $= 9.4 * 10^5$ virions/mL) and varied from very high viral loads ($9.5 * 10^10$ virions/mL) to viral loads near the limit of detection (1.3 virions/mL) (Fig. 2.3A, Fig. 2.6). We next plotted the cumulative distribution of viral load in both populations (Fig. 2.3B). This comparison really represents two extremes: one group is mostly hospitalized, while the other group represents a mostly young and healthy (but infected) college population. Yet, the distributions are extremely

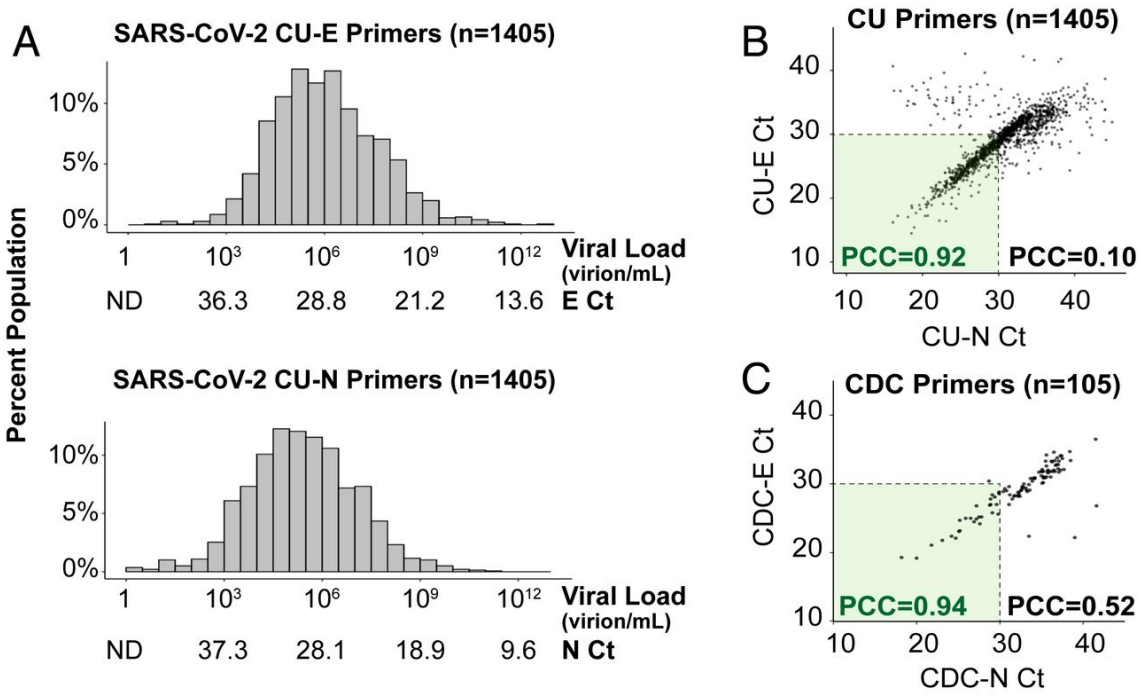


Figure 2.2: **Saliva viral load distribution within our campus population (A)** Distributions are shown of the viral loads measured in the 1,405 positive samples identified on campus during the Fall semester of 2020. Each histogram shows Ct values obtained using TaqMan primer/probe sets targeting either the E gene ("CU-E") or the N gene ("CU-N") of SARS-CoV-2. The horizontal axes are labeled with both the cycle threshold values (Ct) and the corresponding viral loads calculated from the standard curve for each primer set (Fig. 2.1A). ND = no data, as the viral load is below the RT-qPCR detection limit. **(B)** The Ct values resulting from the two primer sets in panel A are highly correlated, especially in samples with high viral loads (Ct value lower than 30). Pearson correlation coefficients (PCC) are shown within and beyond the Ct=30 arbitrary cutoff. **(C)** For 105 of the SARS-CoV-2 positive saliva samples, we ran RT-qPCR side-by-side with 8 different primer sets commonly used in SARS-CoV-2 diagnostic tests (Fig. 2.5). Here, we show the same analysis as in panel B, except with the U.S. Centers for Disease Control's (CDC) primers targeting the E and N genes (see methods).

similar (two-sided two-sample Kolmogorov-Smirnov test, D statistic = 0.03, p-value = 0.97; (Fig. 2.3B)). Therefore, individuals have similar distributions of saliva viral load regardless of symptomatic status, as has also been observed in studies of viral load in anterior nasal or nasopharyngeal swabs [95–100].

2.4.3 A small subset of individuals carries most of the circulating virions

We next analyzed how virus is distributed between individuals within populations. By summing the viral load across individuals based on the interpolated probability density function representing each population, starting with those with the highest viral loads, we find that just 2% of individuals harbor 90% of the circulating virions (Fig. 2.4). This is true in both the university (i.e. asymptomatic) and hospitalized (i.e. symptomatic) populations. Further, 99% of community-circulating virions are accounted for by just 10% of the asymptomatic and 14% of the symptomatic population. In both asymptomatic and symptomatic populations, one single individual with the highest saliva viral load carried more than 5% of the total circulating virions. On the other hand, all individuals with saliva viral loads lower than 10^6 virions per mL combined (representing 50% of the infected individuals) harbor less than 0.02% of the virions in both populations. This can be understood because Ct is a linear representation of logarithmic increases in viral load, so that the viral load increases exponentially as the Ct value decreases (Fig. 2.1). Thus, there is a highly asymmetric distribution of viruses within both populations, with just a small number of people carrying the vast majority of the virus. It remains unknown whether these are special individuals capable of harboring extraordinarily high viral loads, or whether many infected individuals pass through a very short time period of extremely high viral load (see further discussion below). Irrespective of mechanism, it is nevertheless true that, at any given moment in time, a small number of people are harboring the vast majority of virions.

Infectious virions have rarely been isolated from clinical samples of individuals with viral load less than 10^6 virions per mL [99, 101–106]. One hypothesis is that people in this

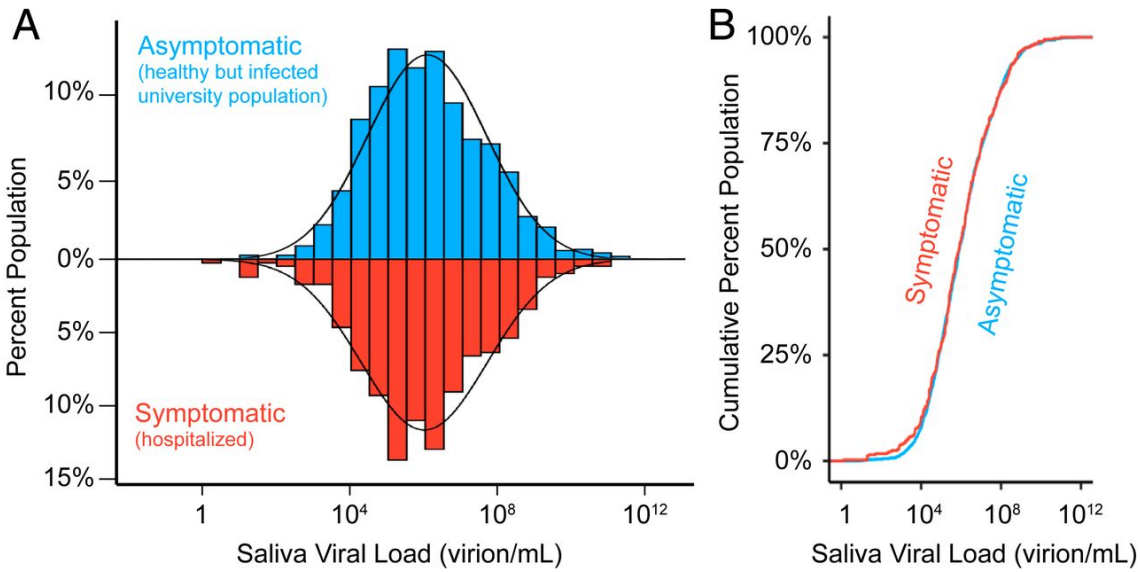


Figure 2.3: **Viral load distributions are similar in asymptomatic and symptomatic populations (A)** A histogram of saliva viral loads in our asymptomatic campus population (N=1405, blue) compared to the same histogram of saliva viral loads from symptomatic (N=404, red) individuals. The latter represents data compiled from the ten studies in Table 2.1). A log-normal probability density function is fitted onto the two distributions given the population mean and standard deviation. **(B)** Empirical cumulative distribution functions (ECDFs) of saliva viral load in the asymptomatic (N=1405, blue) and symptomatic (N=404, red) populations. The similarity of the two ECDFs were assessed with the Kolmogorov-Smirnov test, which resulted in D statistic = 0.03, and p-value = 0.97.

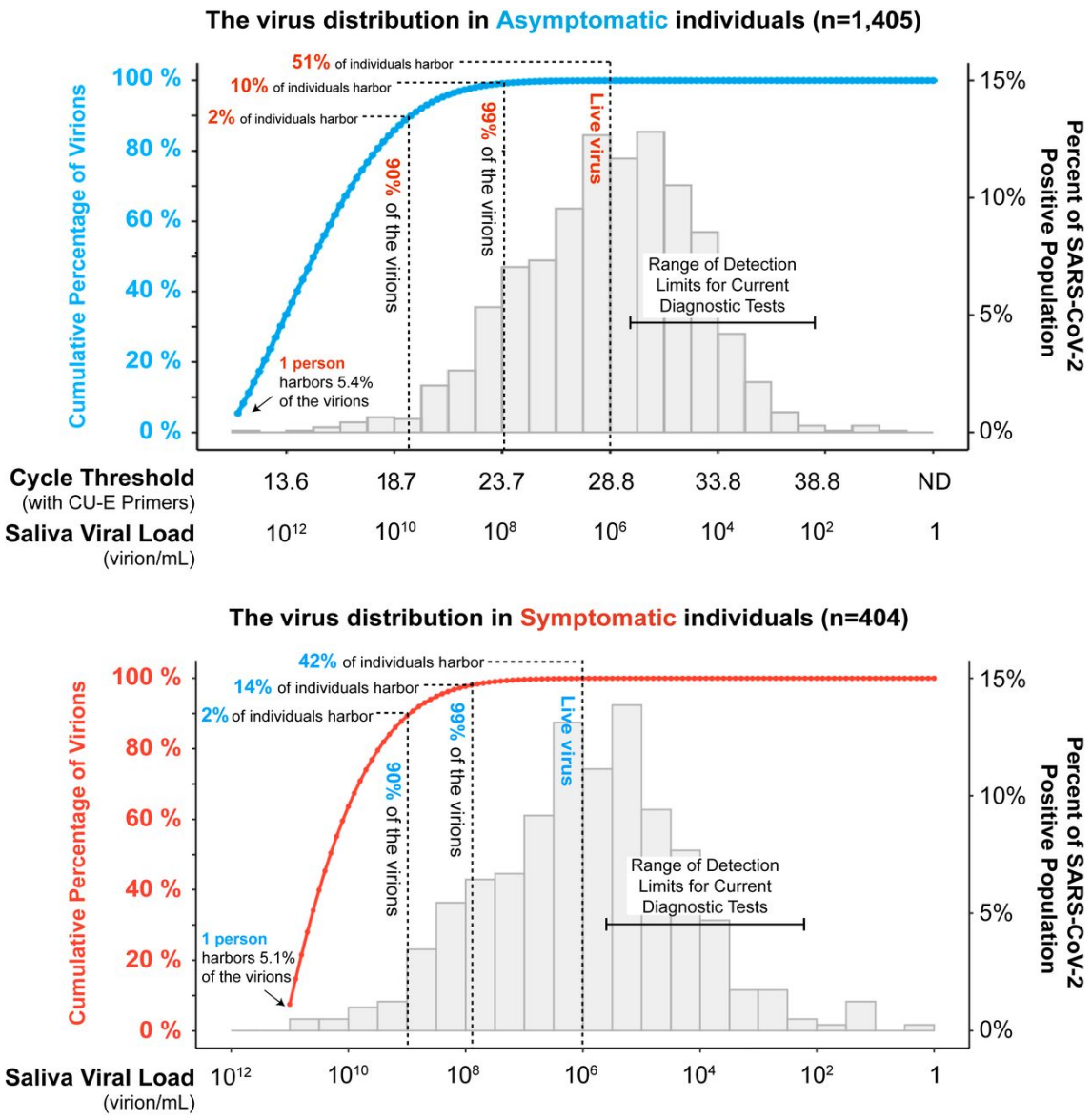


Figure 2.4: **A small percentage of individuals are viral super-carriers** The histograms shown (right Y axes) are the same as were shown in figure 2. Starting from the left of each histogram (i.e. those individuals with the highest viral loads), we calculated the accumulative percentage of total virions as a function of saliva viral load based on the probability density function of the distribution (blue and red lines, and left Y axes). In both asymptomatic (blue line) and symptomatic populations (red line), the portion of population that harbors 90% and 99% of the circulating virus is highlighted by the dashed lines. We estimate that only around 50% (51% and 42% in the panels shown) of individuals who test positive for the virus actually harbor infectious virions, based on the observation that live virus has rarely been isolated from samples with viral loads $< 10^6$ virions per mL [99, 101–106]. For context, the range of detection limits of common SARS-CoV-2 diagnostic testing paradigms (RT-qPCR, antigen testing, and RT-LAMP) are shown. All testing paradigms will capture virtually all infectious individuals and virions, in pre-symptomatic and symptomatic populations alike. Limits of detection are taken from [46, 107, 108].

low range of viral load may simply be shedding viral genomes from damaged tissue that is undergoing repair, and for this reason they may not pose a substantial risk of infecting others. Our distributions suggest that approximately half of the people who test positive may not be infectious to others (Fig. 2.4) based on this line of reasoning.

2.5 Discussion

An important finding herein is that the vast majority of circulating virions in communities are found within the bodies of a small number of individuals. These findings corroborate similar trends observed elsewhere [87–90, 96]. Although it remains to be seen exactly how transmission probability relates to viral load, a strong implication is that these individuals who are viral super-carriers may also be super-spreaders. Higher viral loads have been shown to increase the probability of transmission to others in China [109], in Spain [110], and between pairs of roommates on our University campus [111]. A higher rate of spread by viral super-carriers would be consistent with recent contact tracing analyses suggesting that 80-90% of infections are caused by 10-20% of infected individuals [87–90, 96]. A higher rate of spread by viral super-carriers would also be consistent with the surprisingly low transmission rates being reported between roommates [111], schoolmates [112, 113], and household members [114], which could be explained if only a small fraction of infected individuals have high enough viral loads to facilitate active transmission.

One potential explanation for the differences in viral loads between individuals is that individuals were simply tested at different stages of otherwise similar viral infections. However, longitudinal analyses of individual infections show that peak viral loads vary dramatically between individuals [115–117]. Thus, the parsimonious explanation is that individuals produce different levels of virus. Whether this is due to variation in the immune response, variation in host factors supporting virus replication like ACE2, the specific viral variant infecting, or initial infection site or dose remains to be determined [118–121]. To look at this further, we compared the viral load distributions analyzed herein to a theoretical normal

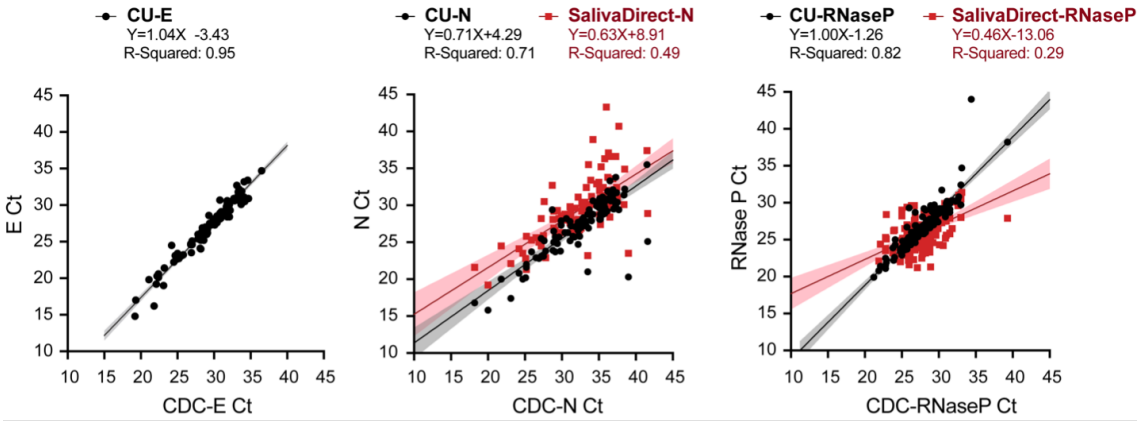


Figure 2.5: **Correlation of Ct values between different primer sets used to quantify saliva viral load.** Using 105 SARS-CoV-2 positive saliva samples, we examined the Ct values obtained with different RT-qPCR multiplex assays and compared them via correlation analysis. For 105 virus-positive saliva samples, 8 different Ct values were generated all in one day from each sample, in a side-by-side direct analysis of the performance of each primer set. Ct values from the Centers for Disease Control primers (CDC-E, CDC-N or CDC-RNaseP) are reported on the X-axes. On the Y-axes are plotted the corresponding Ct values resulted from our university screening primers (CU-E, CU-N or CU-RNaseP) and primer sets used in the SalivaDirect [45] test (SalivaDirect-N and SalivaDirect-RNase P) primer sets.

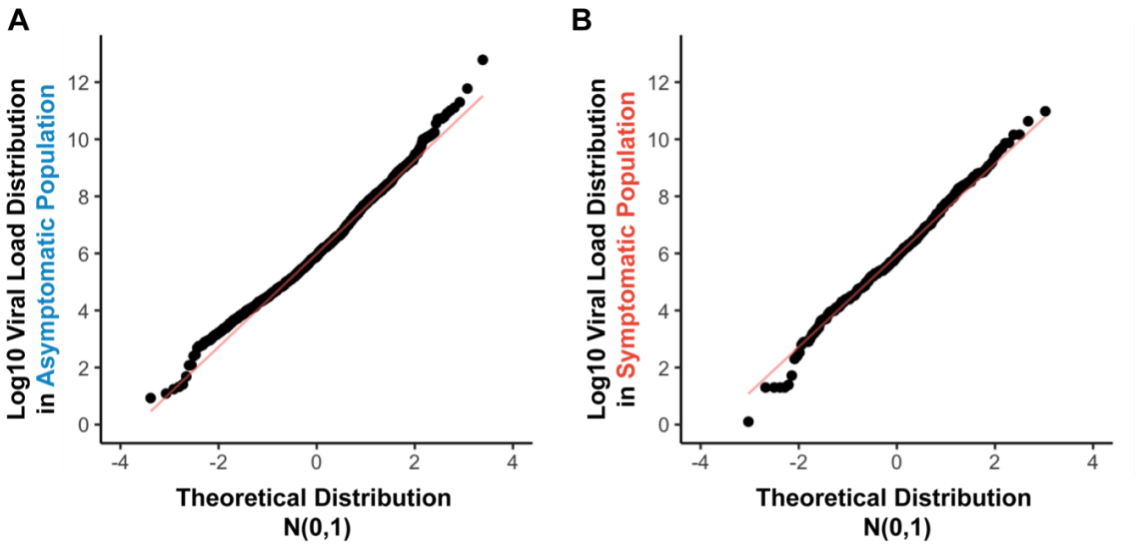


Figure 2.6: **The observed viral loads follow a normal distribution, except at the extreme ends.** We compared the viral load data in each population (Y axis) to the theoretical standard normal distribution (X-axis) using a quantile-quantile plot. The points indicate the empirical quantiles of the datapoints, while the diagonal line (red) indicates the expected quantiles under normal distribution. The data deviates from the Gaussian distribution at the extreme ends, which likely represents individuals with either very high or very low viral loads.

distribution using quantile-quantile plots (Fig. 2.6). The data deviates from the normal distribution at the extreme ends, including in the part of the population with the highest viral loads. This is consistent with the hypothesis that a small percentage of individuals represent a unique population with different capacity for infection than the rest of the population.

The concentration of a majority of the virus in a small fraction of the population at a given time is a critical observation with actionable conclusions. Community screening to identify viral super-carriers within pre-symptomatic and asymptomatic stages of disease will be important, since these individuals will continue to sustain and drive the epidemic if not located. Finding viral super-carriers will have a disproportionately large impact on curbing new COVID-19 infections, yet individuals without symptoms don't tend to seek out testing so screening will need to target healthy populations. Modeling approaches show that one of the most important factors in screening for SARS-CoV-2 will be the speed with which infected people receive their test results (also referred to as turnaround time) [122]. The longer it takes for people to receive their results, the more time goes by where they might unwittingly infect others. Therefore, it is imperative that we find virus super-carriers, and inform them of their infection status in a way that is fast, easy, and accessible. Although detection limits vary between current monitoring and diagnostic paradigms, all are more than capable of finding the majority of infected individuals and the vast majority of circulating virions (Fig. 2.4) [46, 107, 108].

2.6 Methods

2.6.1 Collection of University samples

For sample collection conducted at our university, individuals were asked to fill out a questionnaire (www.colorado.edu/daily-health-form) to confirm that they did not present any symptoms consistent with COVID-19, and to collect no less than 0.5 mL of saliva into a 5 mL screw-top collection tube. Saliva samples were heated at 95°C for 30 minutes on-site

to inactivate the viral particles for safe handling, and then placed on ice or at 4°C before being transported to testing laboratory for quantitative RT-PCR analysis on the same day.

2.6.2 Saliva quantitative RT-PCR used for screening saliva samples on the University of Colorado Boulder campus

For quantitative RT-PCR analysis, the university testing team transferred 75 µL of saliva into one well of a 96-well plate where each well had been pre-loaded with 75 µL 2X TBE buffer supplemented with 1% Tween-20. Of this diluted sample, 5 µL was then added to one well of a separate 96-well plate where each well had been pre-loaded with 15 µL reaction mix composed of: TaqPath 1-step Multiplex Master Mix (Thermo Fisher A28523), nuclease-free water, and triplex primer mix consisting of CU-E, CU-N, and CU-RNaseP primer and probe sets (sequence and concentration specified in the Table 2.2; conditions changed slightly during the semester). The reactions were mixed, spun down, and loaded onto a Bio-Rad CFX96 or CFX384 qPCR machine. Quantitative RT-PCR was run using the standard mode, consisting of a hold stage (25°C for 2 minutes, 50°C for 15 minutes, and 95°C for 2 minutes) followed by 44 cycles of a PCR stage (95°C for 3 seconds, 55°C for 30 seconds, with a 1.6°C/sec ramp up and ramp down rate). Ct values from all campus testing efforts were communicated to us as de-identified data.

2.6.3 Focused analysis of 105 SARS-CoV-2 positive samples

For a smaller subset of 105 samples, as described herein, we did a side-by-side comparison of 3 different quantitative RT-PCR multiplex assays commonly used in SARS-CoV-2 diagnostics. We thawed 105 frozen, de-identified saliva samples which had previously tested positive for SARS-CoV-2 in the campus screening operation and performed all of the following RT-PCR analyses side-by-side on the day of sample thawing.

First, 25 µL of thawed, previously heat-treated saliva was transferred into one well of a 96-well plate where each well had been pre-loaded with 25 µL 2xTBE buffer supplemented

with 1% Tween-20. Next, 5 µL of the diluted sample was added to separate 96-well plates where each well had been pre-loaded with 15 µL reaction mix composed of: TaqPath 1-step Multiplex Master Mix (Thermo Fisher A28523), nuclease-free water and CDC triplex primer mix or CU triplex primer mix (sequence and concentration specified in the Table 2.2). The reactions were mixed, spun down, and loaded onto a Bio-Rad CFX96 qPCR machine. Quantitative RT-PCR was run using the standard mode, consisting of a hold stage (25°C for 2 minutes, 50°C for 15 minutes, and 95°C for 2 minutes) followed by 44 cycles of a PCR stage (95°C for 3 seconds, 55°C for 30 seconds, with a 1.6°C/sec ramp up and ramp down rate). Each plate also contained two wells of negative control template (5 µL nuclease-free water diluted 1:1 with 2XTBE supplemented with 1% Tween-20) and two wells of positive control template (5 µL synthetic SARS-CoV-2 RNA (Twist Biosciences 102024) diluted to 1000 genome copies/µL, and 5 µL total human reference RNA (Agilent 750500) diluted to 10ng/µL in nuclease-free water).

We also performed the SalivaDirect TaqMan RT-qPCR analysis [45] on each of these samples. 75 µL of each saliva specimen was combined with 9.4 µL of Proteinase K (20 mg/mL, NEB, P8107S). Samples were incubated at ambient temperature for 15 minutes and then heated to 95°C for 5 minutes to inactive the Proteinase K. Next, 5 µL of saliva was used as template in a 20 µL reaction that also contained 1X TaqPath 1-step Multiplex Master Mix, nuclease free water and primer and probe sets at concentrations described below. The RT-qPCR was run on the BioRad CFX96 qPCR machine using the same program described for the CU assays [45].

Table 2.1: **Studies from which viral loads in symptomatic individuals were derived** (All studies indicated that saliva samples were self-collected from COVID-19 patients at the indicated locations. In all cases, the authors reported virus concentrations in the original saliva sample.)

Study	Hospital, Country	Gender (M/F)	Age	Method
To et al.[49]	Princess Margaret Hospital and Queen Mary Hospital, Hong Kong	13/10	37-75	Posterior oropharyngeal saliva followed by RT-qPCR
To et al.[123]	Princess Margaret Hospital and Queen Mary Hospital, Hong Kong	7/5	37-75	Posterior oropharyngeal saliva followed by RT-qPCR
Zhang et al.[124]	Wuhan Pulmonary Hospital, China	NA	NA	Saliva from oral swab followed by RT-qPCR
Hanege et al.[125]	Goztepe Education and Research Hospital, Turkey	11/18	26-70	Self-collected saliva followed by RT-qPCR
Procop et al.[126]	Cleveland Clinic, USA	25/14	18-82	Self-collected saliva followed by RT-qPCR
Zheng et al.[124]	First Affiliated Hospital, College of Medicine, China	58/38	44-64	Self-collected saliva after deep cough followed by RT-qPCR
Yoon et al.[127]	Korea University Guro Hospital, Korea	0/2	46-65	Self-collected saliva followed by RT-qPCR
Wyllie et al.[91]	Yale New Haven Hospital, USA	NA	NA	Self-collected saliva followed by RT-qPCR
Yokota et al.[128]	Hokkaido University Hospital, Japan	25/17	27-93	Self-collected saliva followed by RT-qPCR
Zhu et al.[129]	Central Hospital of Xiangtan, China	16/16	34-54	Self-collected saliva followed by RT-qPCR

Table 2.2: **The qRT-PCR TaqMan primer/probe sets used for university screening and focused analysis.** *Explanation for some of the TaqMan fluorophores and quenchers used: IAbRQSp, Iowa Black Dark Quenchers RQ; IABkFQ, Iowa Black Dark Quenchers FQ; ZEN, internal quencher; TexRd, Texas Red; HEX, Hexachloro-fluorescein; FAM, fluorescein.

Assay name	Primer/probe set target and designation	Primer or probe name	1× concentration, nM	Sequence (5′ to 3′)*
CU	SARS-CoV-2 E gene "CU-E"	E_Sarbeco_F1 (IDT 10006888)	400	ACAGGTACGTTAATAGTTAATAGCGT
		E_Sarbeco_R2(IDT 10006890)	400	ATATTGCAGCAGTACGCACACA
		E_Sarbeco_P(IDT Custom)	200	TexRd-ACACTAGCCATCCTTACTGCGCTTCG-IAbRQSp
	SARS-CoV-2 N gene "CU-N"	nCOV_N1_F (IDT 10006830)	500	GACCCCAAAATCAGCGAAAT
		nCOV_N1_R (IDT 10006831)	500	TCTGGTTACTGCCAGTTGAATCTG
		nCOV_N1_P (IDT Custom)	250	HEX-ACCCCGCAT-ZEN-TACGTTTGGTGGACC-IABkFQ
	Human RNase P"CU-RNaseP"	RNaseP_F (IDT 10006836)	50	AGATTTGGACCTGCGAGCG
		RNaseP_R (IDT 10006837)	50	GAGCGGCTGTCTCCACAA GT
		RNase_P_P (IDT 10006838)	50	FAM-TTCTGACCT-ZEN-GAAGGCTCTGCGCG-IABkFQ
CDC	SARS-CoV-2 E gene "CDC-E"	E_Sarbeco_F1 (IDT 10006888)	400	ACAGGTACGTTAATAGTTAATAGCGT
		E_Sarbeco_R2 (IDT 10006890)	400	ATATTGCAGCAGTACGCACACA
		E_Sarbeco_P (IDT 10006893)	200	FAM-ACACTAGCCATCCTTACTGCGCTTCG- IABkFQ
	SARS-CoV-2 N gene "CDC-N"	nCOV_N2_F (IDT 10006833)	600	TTACAAACATTGGCCGCAAA
		nCOV_N2_R (IDT 10006834)	600	GCGCGACATTCCGAAGAA
		nCOV_N2_P (IDT 10007049)	300	SUN- ACAATTTGCCCCCAGCGCTTCAG - IABkFQ
	Human RNase P "CDC-RNaseP"	RNaseP_F (IDT 10006836)	50	AGATTTGGACCTGCGAGCG
		RNaseP_R (IDT 10006837)	50	GAGCGGCTGTCTCCACAA GT
		RNase_P_P (IDT 10007062)	50	ATTO647-TTCTGACCTGAAGGCTCTGCGCG - IABkFQ
SalivaDirect	SARS-CoV-2 N gene "SalivaDirect-N"	nCOV_N1_F (IDT 10006830)	400	GACCCCAAAATCAGCGAAAT
		nCOV_N1_R (IDT 10006831)	400	TCTGGTTACTGCCAGTTGAATCTG
		nCOV_N1_P (IDT 10006832)	200	FAM-ACCCCGCATTACGTTTGGTGGACC-IABkFQ
	Human RNase P "SalivaDirect-RNaseP"	RNaseP_F (IDT 10006836)	150	AGATTTGGACCTGCGAGCG
		RNaseP_R (IDT 10006837)	150	GAGCGGCTGTCTCCACAA GT
		RNase_P_P (IDT 10007062)	200	ATTO647-TTCTGACCTGAAGGCTCTGCGCG - IABkFQ

Chapter 3

Saliva TwoStep for rapid detection of asymptomatic SARS-CoV-2 carriers

---

Adapted from: Yang Q, Meyerson NR, Clark SK, Paige CL, Fattor WT, Gilchrist AR, Barbachano-Guerrero A, Healy BG, Worden-Sapper ER, Wu SS, Muhlrad D, Decker CJ, Saldi TK, Lasda E, Gonzales P, Fink MR, Tat KL, Hager CR, Davis JC, Ozeroff CD, Brisson GR, McQueen MB, Leinwand LA, Parker R, Sawyer SL. Saliva TwoStep for rapid detection of asymptomatic SARS-CoV-2 carriers. eLife. 2021 Mar 29;10:e65113.

3.1 Contributions

This study is the fruit of a combined effort of numerous scientists who worked together in containing the spread of SARS-CoV-2 on campus. Back in March, 2020, when everything was shutdown, we were granted essential worker status and tasked to come up with solutions for an on-campus infection screening test. Due to the supply chain bottleneck, it became increasingly difficult to order basic laboratory supplies, and especially the supplies for nasal-swab based SARS-CoV-2 RT-qPCR test. In response, we came up with this saliva-based rapid SARS-CoV-2 screening test based on isothermal amplification technology. Throughout this process, I worked alongside with my colleagues, especially with Dr. Nick Meyerson in designing, optimizing and validating the assay. I was responsible for designing the primers, as well as optimizing the reaction condition for this screening test. In addition, with the help from the university COVID-19 response team, I was hence able to carry out the validation study that used hundreds of samples collected on campus. Eventually, I generated the figures

and wrote the manuscript with guidance from Dr. Meyerson and my advisor Dr. Sawyer. Together, none of this would be possible without all of the selfless scientists who helped mitigate the SARS-CoV-2 outbreak on campus during 2020-2021 academic years, and they are listed here as the co-authors of this manuscript.

3.2 Abstract

Here, we develop a simple molecular test for SARS-CoV-2 in saliva based on reverse transcription loop-mediated isothermal amplification (RT-LAMP). The test has two steps: 1) heat saliva with a stabilization solution, and 2) detect virus by incubating with a primer/enzyme mix. After incubation, saliva samples containing the SARS-CoV-2 genome turn bright yellow. Because this test is pH dependent, it can react falsely to some naturally acidic saliva samples. We report unique saliva stabilization protocols that rendered 295 healthy saliva samples compatible with the test, producing zero false positives. We also evaluated the test on 278 saliva samples from individuals who were infected with SARS-CoV-2 but had no symptoms at the time of saliva collection, and from 54 matched pairs of saliva and anterior nasal samples from infected individuals. The Saliva TwoStep test described herein identified infections with 94% sensitivity and >99% specificity in individuals with sub-clinical (asymptomatic or pre-symptomatic) infections.

3.3 Introduction

Disease screening is one of the most basic and powerful tools in the public health arsenal. Screening tests identify unknown illness in apparently healthy or asymptomatic individuals. In the case of dangerous pathogens, screening tests serve to direct potential carriers of the pathogen into the healthcare system for confirmatory testing, and to alert them that they could possibly infect others while they await confirmatory results. If dangerous pathogens are spreading at high rates, individuals will need to be screened frequently. As such, screening tests should operate with minimal requirements for laboratory equipment

and labor, such that they are community-deployable and don't burden the critical pipelines for diagnostics. In the current SARS-CoV-2 pandemic, body temperature is a ubiquitous screening test being used on apparently-healthy people around the world. However, using elevated body temperature as a sign of SARS-CoV-2 infection lacks specificity for this particular pathogen and sensitivity in identifying asymptomatic carriers [130]. To help fill in the need for more reliable screening tests, here we present a simple and portable assay that detects the SARS-CoV-2 genome in saliva with specificity and sensitivity, even in samples from individuals with no symptoms at the time of saliva collection.

LAMP (loop-mediated isothermal amplification) is a simple nucleic acid diagnostic concept that has existed for more than 20 years [131]. It has been used in diverse and even remote settings to test samples for the presence of viral nucleic acids [132, 133]. LAMP utilizes loop forming primers and strand-displacement polymerases to achieve isothermal amplification of a target nucleic acid template, and therefore does not require a thermal cycler. LAMP assays can be performed anywhere because they simply require pipettors and a heating source (e.g. water baths or heat blocks) as equipment [132]. LAMP assays offer robust amplification of target material and can produce on the order of 109 copies of the target in an hour-long reaction [131]. Successful amplification in LAMP reactions can be directly visualized by simply looking at the reaction tube, where the reaction mix changes color upon successful target amplification. These colorimetric changes can be triggered by pH indicator dyes or metal ion indicators, which change color when successful target amplification changes the chemistry within the reaction tube [131, 134, 135]. If more sophisticated visualization equipment is available, other indicators can used. Intercalating fluorescent DNA dyes or quenched fluorescent probes can be used which emit fluorescent signal over time during amplification [136, 137]. Alternately, real-time measurements of turbidity in the tube can be used to measure changes in turbidity resulting from magnesium pyrophosphate formation as amplification proceeds [138]. RT-LAMP (reverse transcription - loop-mediated isothermal amplification), where a reverse transcription step is added upstream of the LAMP reaction,

adapts all of these protocols for detection of RNA. RT-LAMP with a simple visual color change that occurs in sample tubes containing SARS-CoV-2 could be well suited as a rapid and deployable community-based screening test [139].

Recent studies have shown that saliva has high diagnostic value for SARS-CoV-2 [45, 91, 92, 140, 141]. Compared to nasopharyngeal swabs, saliva samples harbor similar levels of viral load while being easier to obtain via self-collection. Several groups have developed RT-LAMP tests to detect SARS-CoV-2 in saliva samples [141–148]. However, due to pH variability between saliva samples, RT-LAMP often has a high rate of false positives when used with the common pH-dependent dye phenol red [136, 142]. In RT-LAMP reactions containing phenol red, reactions start as pink/red but turn strongly yellow at pH 6.8 and below. When RT-LAMP amplifies a target, hydrogen ions are released during dNTP incorporation. This causes a drop in pH within the tube to pH 6.0 – 6.5, triggering the color change to yellow [135]. Human saliva naturally varies in pH between 6.8 and 7.4 [149], posing a significant problem in this pH-dependent assay. In fact, we find about 7% of human saliva samples are naturally acidic enough to immediately trigger phenol red-containing reactions to change to yellow without any target amplification (Figure 3.1A Left). If acidic samples are not anticipated and managed, colorimetric RT-LAMP has the potential to produce a high false-positive rate.

Here, we combine the simplicity of RT-LAMP and the non-invasive nature of saliva to develop an effective screening test for SARS-CoV-2. This test does not require RNA purification but rather works directly with human saliva. We optimized a saliva stabilization solution that 1) neutralizes the variability of human saliva and essentially eliminates false positives, 2) lowers the viscosity of saliva, and 3) stabilizes RNA for analysis in the test. We validated the RT-LAMP test using a large cohort of saliva and matched nasal swab specimens collected from our local university population, comparing the test to two other quantitative RT-PCR-based SARS-CoV-2 tests (one nasal test and one saliva test). We found our optimized RT-LAMP procedure performs consistently with high specificity and

sensitivity, even though our samples were largely from individuals who had no reported symptoms at the time of sample collection. Based on our experience performing screening on our university campus and elsewhere, we provide in the appendix extensive operational details and recommendations for successful community deployment of this SARS-CoV-2 screening test.

3.4 Results

3.4.1 Optimized universal saliva stabilization conditions for RT-LAMP

To deal with the variability in pH of human saliva, we optimized a basic saliva stabilization solution by titrating in various concentrations of sodium hydroxide (NaOH). We performed this optimization using a control RT-LAMP primer set, "RNaseP," which amplifies the mRNA transcript produced from the human POP7 gene (primer set developed previously [150]). Our goal was to increase the pH of all saliva samples well above the indicator flip-point of pH 6.8, while not making the samples so basic that they couldn't reach this pH upon successful target amplification. We found that human saliva containing 14.5 mM NaOH is optimal to inhibit false positives caused by saliva acidity (N=96; Figure 3.1A, right) without impeding the intended color change during amplification (Figure 3.1B). In addition, we designed our saliva stabilization solution to also include a chelating agent (1 mM EDTA final concentration) and Proteinase K to inhibit RNases, both of which help preserve virion RNA and therefore to increase sensitivity (note that Proteinase K will inhibit the RT-LAMP reaction if it does not go through a heat inactivation step prior to that reaction). Finally, the saliva stabilization solution contains TCEP, which aids in RNA stabilization by breaking disulfide bonds present in RNases and helping to reduce saliva viscosity. Our optimized saliva stabilization solution (2X solution: 5 mM TCEP, 2 mM EDTA, 29 mM NaOH, 100 μg/mL Proteinase K, diluted in DEPC-treated water) is key to this test. For additional advice on controlling the acidity of reactions see the appendix.

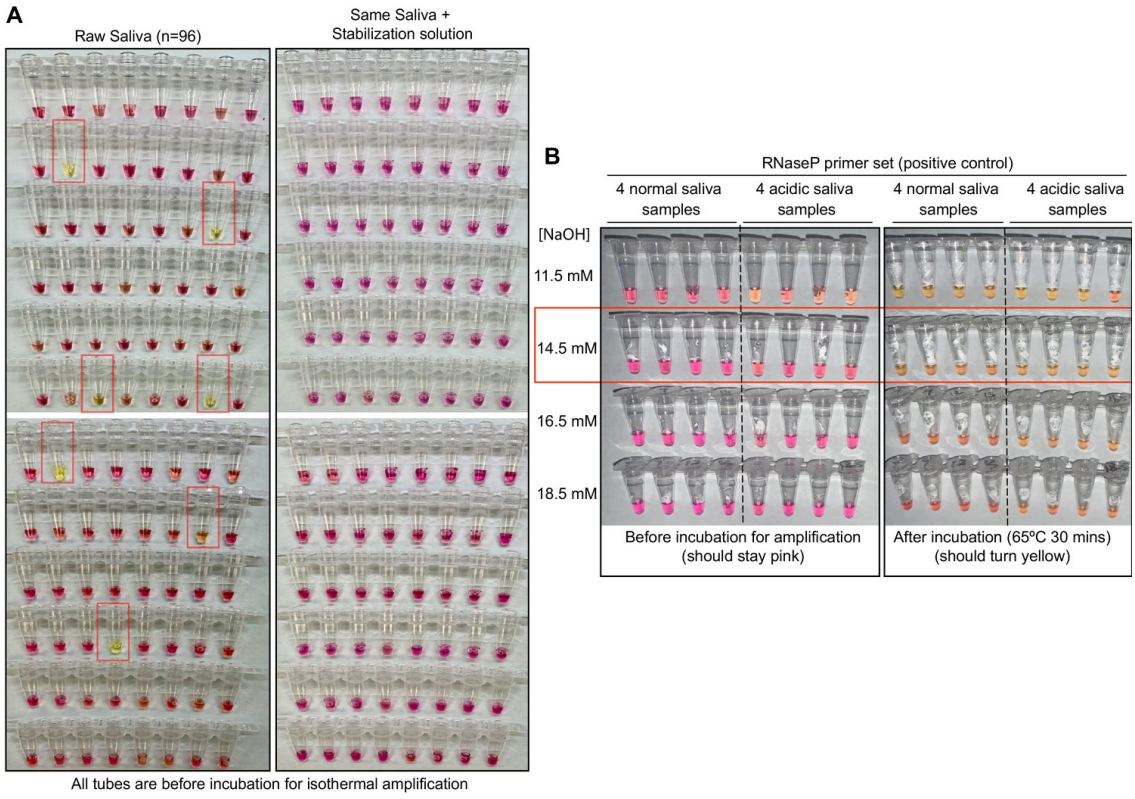


Figure 3.1: **Optimized strategy for controlling natural variability in saliva pH. (A)** Here, saliva samples from 96 different individuals are analyzed for the prevalence of natural acidity extreme enough to trigger the pink-to-yellow color change of phenol red even before isothermal amplification. Each saliva sample was combined 1:1 with water (left) or 2X saliva stabilization solution (right; methods) and heated at 95ºC for 10 minutes to liberate RNA from virions. 2 µL of each was then added to 18 µL RT-LAMP reaction mix (2X Colorimetric RT-LAMP Master Mix, RNase P primers, nuclease-free water). The pictures show tubes immediately after samples and master mix are combined, before any incubation steps are undertaken to commence isothermal amplification. With raw saliva, 7 out of 96 tubes turned yellow at this step (highlighted in red boxes). These are false-positives, because no amplification reaction has occurred. None of these 96 saliva samples mixed with saliva stabilization solution turned the reaction tube prematurely yellow. **(B)** Here, we show the method that we had used to identify the ideal pH of the saliva stabilization solution used in panel A and throughout this paper. We chose 4 normal and 4 acidic saliva samples and mixed each 1:1 with 2X saliva stabilization solution containing NaOH at various concentrations (final molarity of NaOH after mixing shown). Samples were then heated at 95ºC for 10 minutes and combined with RT-LAMP reaction mix and control primers recognizing the human RNase P transcript. Before incubation, all tubes should be pink, and after incubation all tubes should be yellow. Based on this, the red box indicates the final optimal NaOH concentration chosen.

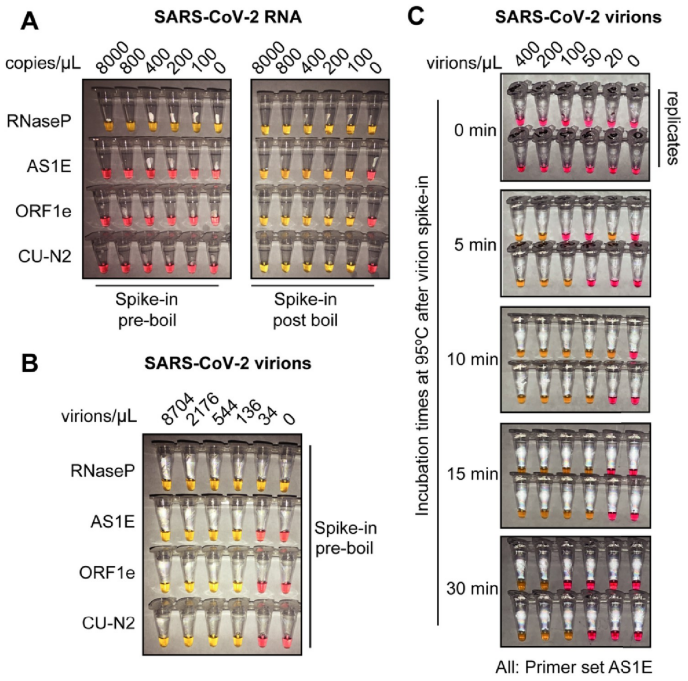


Figure 3.2: **Optimized heat inactivation for safely detecting SARS-CoV-2 in human saliva. (A)** This experiment shows that heating at 95°C for 10 minutes degrades viral RNA when it is not in the form of virions. Saliva samples were diluted 1:1 with saliva stabilization solution. *In vitro* transcribed SASR-CoV-2 RNA was spiked into the diluted saliva to reach the indicated concentrations before (left) or after (right) the heating at 95°C for 10 minutes. To match other experiments, the indicated concentration represents the copies of SASR-CoV-2 RNA in the original undiluted saliva. The samples were then subjected to RT-LAMP at 65°C for 30 minutes. In this colorimetric version of RT-LAMP, reactions remain pink when no amplification occurred, and turned yellow if there was an amplification event. An RT-LAMP primer set targeting the human RNaseP transcript is included as a host RNA amplification control in addition to the three SARS-CoV-2 primer sets shown in panel A. **(B)** This experiment shows that heating saliva at 95°C for 10 minutes does not degrades viral RNA when it is in the form of virions. Saliva samples were spiked with the indicated concentrations of heat-inactivated SARS-CoV-2 virions before being diluted 1:1 with saliva stabilization solution. Samples were then heated at 95°C for 10 minutes and subjected to RT-LAMP similarly to the experiment shown in panel A. **(C)** Results illustrate the optimal incubation time at 95°C to liberate SARS-CoV-2 RNA from virions. Saliva samples were spiked with the indicated concentrations of heat-inactivated SARS-CoV-2 virions before being diluted 1:1 with saliva stabilization solution. Samples were then heated at 95°C for the indicated amount of time, and subjected to RT-LAMP similarly to the experiment shown in panel B. Without heating, no SARS-CoV-2 RNA can be detected with RT-LAMP, presumably because virions remain intact and the viral RNA is not accessible by the amplification enzymes. Amplification is somewhat inconsistent at 5 and 30 minutes possibly because at 5 minutes hardly any RNA has been liberated, and by 30 minutes it has been largely degraded. However, 10 or 15 minutes at 95°C appears to provide just the right balance between liberating and preserving RNA. All reactions contain the AS1E primer set. Duplicates are presented at each time point.

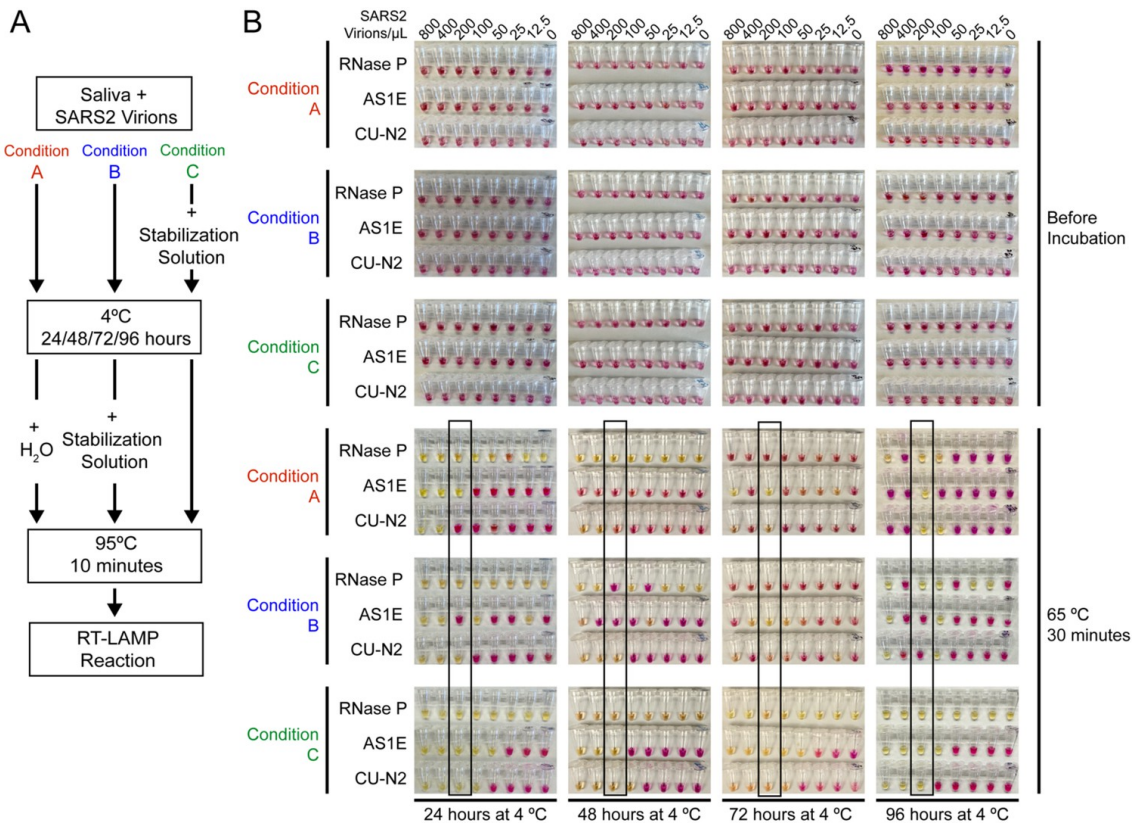


Figure 3.3: **Saliva samples are stable at 4ºC for at least 4 days before processing, if stored in saliva stabilization solution.** **(A)** Schematic of the experimental conditions. **(B)** RT-LAMP reaction result before and after the isothermal amplification. Saliva samples were spiked with heat-inactivated SARS-CoV-2 virions at the indicated concentration and mixed 1:1 with saliva stabilization solution or nuclease-free water before/after storing at 4ºC for 24, 48, 72 and 96 hours. Samples were then heated at 95ºC for 10 minutes and analyzed using RT-LAMP with the indicated primer sets. Condition C, which is the condition used in our test, performs robustly and is sensitive to the limit of detection even after 96 hours storage at 4ºC. The stated limit of detection of 200 virions/μL is boxed.

3.4.2 Optimized RT-LAMP primer sets for detecting SARS-CoV-2 in human saliva

A critical parameter in RT-LAMP is primer design because RT-LAMP requires 4-6 primers all working together [131]. We found that the "AS1E" set, developed by Rabe et. al. and targeting the ORF1ab region of the SARS-CoV-2 genome, performs very well [147]. However, in order to target two distinct regions from the SARS-CoV-2 genome, we designed and tested a large number of additional primer sets. Two of our custom sets, "ORF1e" targeting the virus ORF1ab gene, and "CU-N2" targeting the virus N gene, exhibited similar sensitivity and amplification efficiency as the AS1E set, as determined using real-time fluorescence monitoring of RT-LAMP products (Figure 3.4A, 3.5). We next confirmed that these primer sets were both compatible with saliva preserved in our saliva stabilization solution and with colorimetric RT-LAMP (Figure 3.4B).

3.4.3 Addressing biosafety concerns through heat inactivation

Next, we addressed the biosafety concerns of handling potentially infectious saliva samples. Recent studies suggest that incubation for 3 minutes at 95^{o}C is sufficient to inactivate SARS-CoV-2 virions [151]. However, when heating saliva samples for downstream analysis of RNA, one must balance heating long enough to liberate the target RNA from virions with not heating for so long that the target RNA will be degraded. Heating at 95 ^{o}C does degrade SARS-2-CoV RNA that is spiked directly into saliva samples but does not degrade viral RNA when it is spiked into samples within SARS-CoV-2 virions (Figure 3.2). A 10-minute incubation of saliva samples at 95^{o}C was found to be optimal (Figure 3.2). We designed our test procedure such that testing personnel avoid handling open tubes until after this step to increase biosafety (Appendix).

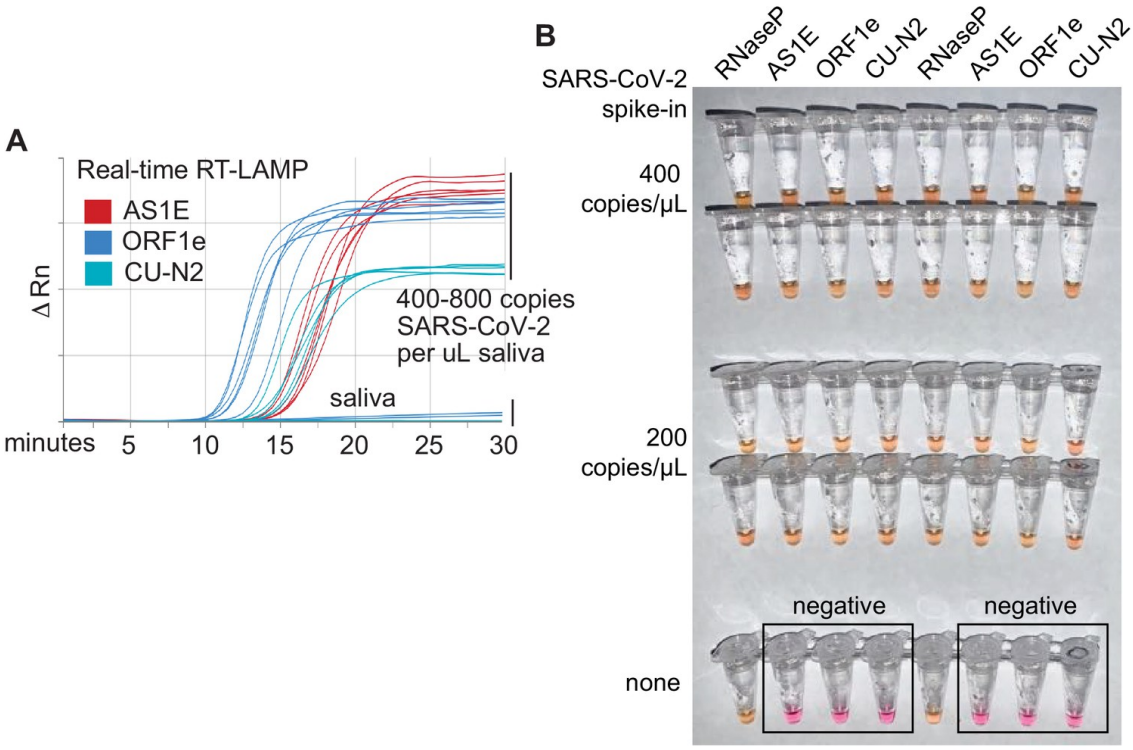


Figure 3.4: **Optimized RT-LAMP primer sets for detecting SARS-CoV-2 in human saliva. (A)** Three RT-LAMP primer sets targeting the SARS-CoV-2 genome (AS1E [147], ORF1e, and CU-N2) were tested with real-time RT-LAMP. Saliva was mixed 1:1 with 2X saliva stabilization solution, heated at 95°C for 10 minutes, and then spiked with *in vitro* transcribed SARS-CoV-2 RNA at the indicated concentrations. 4 μL of this was added to a master mix containing primers and NEB's WarmStart LAMP 2x Master Mix in a final reaction volume of 20 μL. Reactions were incubated at 65°C and a fluorescence reading was taken every 30 seconds. EvaGreen was used to monitor amplification products in real-time (X-axis) using a QuantStudio3 quantitative PCR machine. There are 9 lines for each of the three primer sets because three concentrations of spiked in SARS-CoV-2 RNA were each tested in triplicate (0, 400, 800 copies / μL saliva). The saliva samples without SARS-CoV-2 RNA spike in are shown as flat lines. When concentrations are given herein, denominator refers to the raw, pre-diluted saliva sample. The normalized change in fluorescence signal (ΔRn) is shown on the Y-axis. B) Saliva mixed 1:1 with 2X saliva stabilization solution was heated (95°C for 10 minutes) and then spiked with SARS-CoV-2 RNA at the indicated concentrations. Replicates were tested by RT-LAMP with the control RNaseP primer set and three distinct SARS-CoV-2 primer sets (AS1E, ORF1e, and CU-N2). All samples scored positive except those boxed, which are saliva samples that contain no SARS-CoV-2 RNA, as expected.

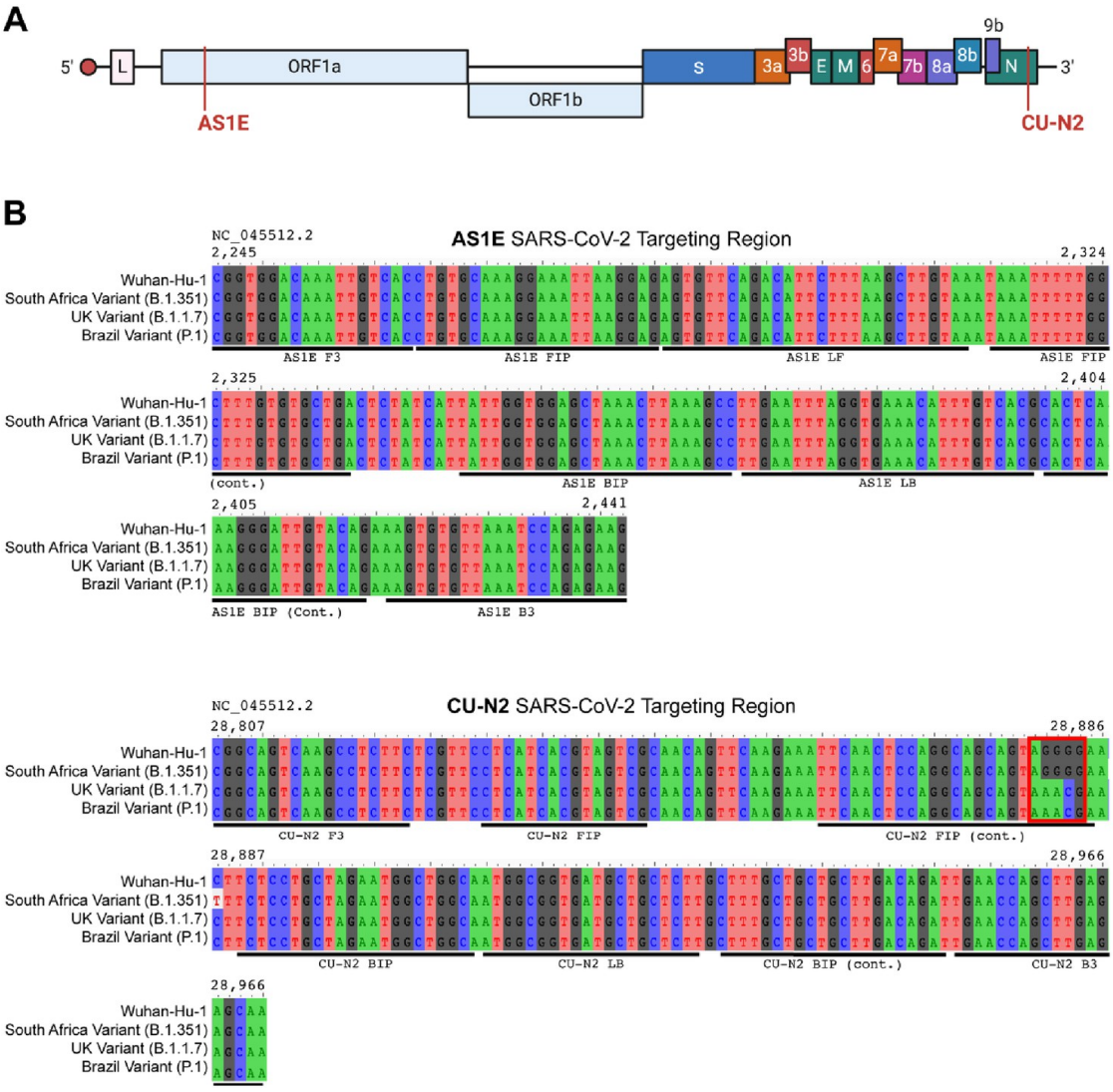


Figure 3.5: **Saliva TwoStep primers will detect most or all currently circulating viral variants of concern. (A)** Genome map of SARS-CoV-2 with the regions targeted RT-LAMP primers highlighted in red. SARS-CoV-2 genome map is adapted from BioRender. B) Sequence alignments of regions of the key SARS-CoV-2 genome variants targeted by RT-LAMP primer sets AS1E and CU-N2. Binding regions of each individual primer set component is highlighted in underlying horizontal bars. The SARS-CoV-2 genome region targeted by AS1E primer set is conserved among all variants. For CU-N2, the red box highlights region of sequence variation that might render CU-N2 primer set less effective to identify the UK and Brazil variants. The coordinate of the genome sequence is based on the SARS-CoV-2 reference genome (NCBI NC_045512.2). The SARS-CoV-2 variant representative genomes are downloaded from GISAID (South Africa Variant B.1.351: hCoV-19/South Africa/KRISP-EC-K004572/2020; UK Variant B.1.1.7: hCoV-19/England/MILK-9E2FE0/2020; Brazil Variant P.1: hCoV-19/USA/VA-DCLS-2185/2020).

3.4.4 Assessment of sample stability during storage

Stability of saliva samples from the time of collection to the time of processing and analysis is important if testing cannot be performed immediately, or if the tests are being conducted in batches. Saliva samples containing purified virions and diluted with 2X saliva stabilization solution were stored at 4^{o}C for 24, 48, 72, or 96 hours before being inactivated at 95^{o}C and analyzed using colorimetric RT-LAMP (Figure 3.3). We tested saliva collection and storage over a range of SARS-CoV-2 virion spike-in concentrations. We observed no significant changes in sample stability and the test detection limit over this time course, suggesting that saliva samples stored in saliva stabilization solution at 4^{o}C are stable for at least four days.

3.4.5 Determining the limit of detection

We next sought to carefully evaluate the limit of detection for this test. The lowest concentration at which positive samples were reliably identified was 200 virions/μL in saliva (red box, summary table in Figure 3.6A). We next tested 20 replicates at this concentration (200 virions/μL) using all four primer sets (Figure 3.6B). The ORF1e primer set was not consistent in its performance at 200 virions/μL. Therefore, we decided to eliminate the ORF1e primer set from our testing panel and define a final colorimetric RT-LAMP test that includes primer sets RNaseP, AS1E, and CU-N2. Note that the limit of detection refers to the virus concentration that can be identified $> 95\%$ of the time, and the assay does often detect the virus at even lower concentrations.

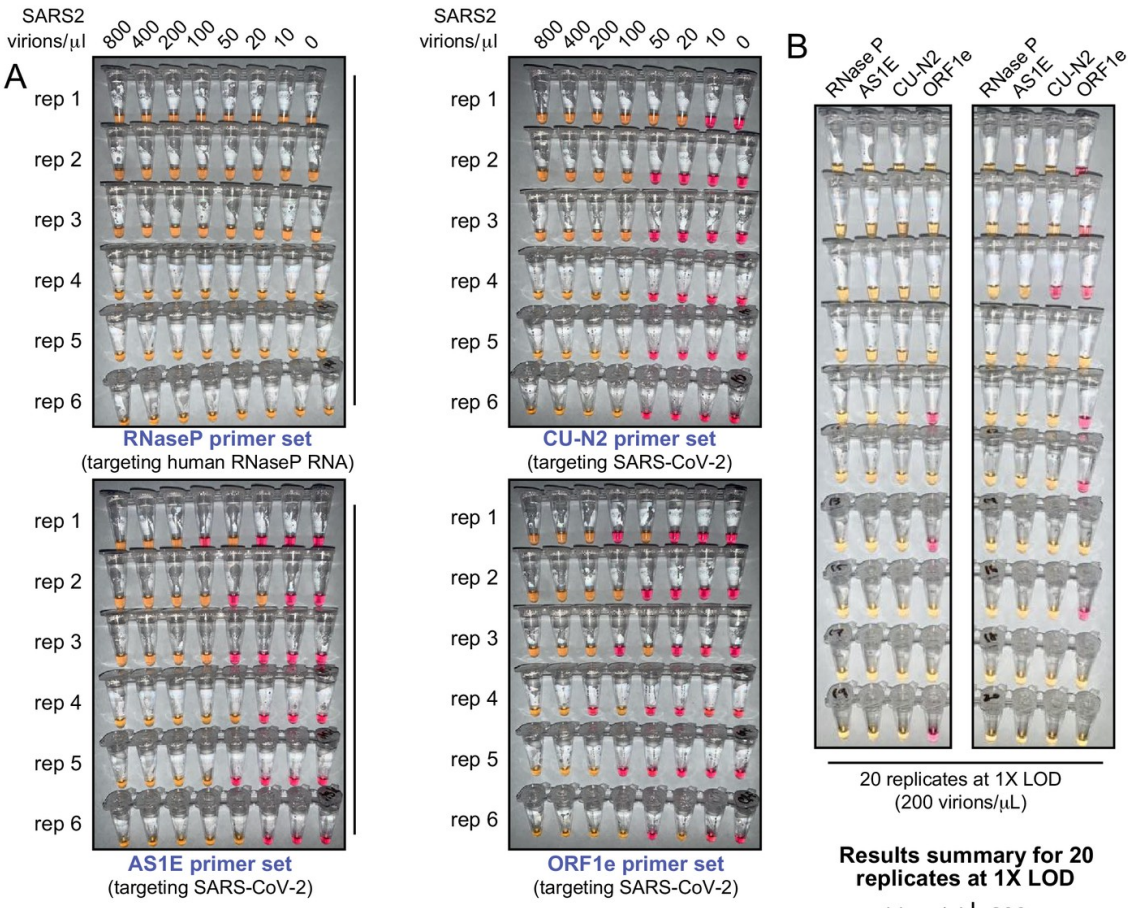
We considered that contaminants in saliva and/or components of the saliva stabilization solution might be suppressing the overall RT-LAMP reaction efficiency by acting in inhibitory ways. On the contrary, we found that when synthetic SARS-CoV-2 RNA is directly added to the RT-LAMP reaction mix (in the absence of saliva and the stabilization solution), we were unable to achieve a better detection limit lower than 200 genome copies/μL (Figure

3.7). This suggests the observed detection limit represents the upper performance limit of the presented RT-LAMP assay, and the saliva and stabilization solution have little to no negative impact to the test performance. In fact, multiple observations suggest that RNA degradation is observed in the absence of stabilization solution, resulting in less consistent testing results (Figure 3.1A; Figure 3.3; Figure 3.7).

We next performed a blinded study. Heat-inactivated virions were spiked into human saliva at various concentrations at or above the limit of detection (200 virions/μL), and these as well as uninfected saliva samples were blinded and passed to a second member of our personnel. After running the RT-LAMP test on 60 such samples, only one positive sample scored as inconclusive. In that sample the SARS-CoV-2 primer set CU-N2 failed, while the other primer set detecting SARS-CoV-2 correctly identified the sample (Figure 3.8). All negative samples were scored correctly (100% specificity, binomial 95% confidence interval [88%,100%]). Conservatively counting the inconclusive test as a false negative lead to a sensitivity estimate of 97% (binomial 95% confidence interval [93%,100%]). See Figure 3.9C for a breakdown by primer set.

3.4.6 Evaluation on human samples

SARS-CoV-2 screening was initiated on the University of Colorado Boulder campus starting in the summer/fall of 2020. Saliva samples were taken weekly from residents of dormitories and at several testing sites throughout the campus. Participants were asked to refrain from eating or drinking 30 minutes prior to sample collection, and to not participate if they were experiencing any symptoms consistent with COVID-19. These individuals were either pre-symptomatic at the time of saliva collection, or they never developed symptoms throughout the course of infection (we don't have the necessary follow-up data to delineate these two outcomes). All saliva samples were first analyzed by a quantitative RT-PCR method performed directly on saliva mixed 1:1 with 2X TBE buffer containing 1% Tween-20 [94]. An optimized multiplex quantitative RT-PCR reaction was used targeting the E and



Results summary for LOD curves (6 replicates)

virions/µl	800	400	200	100	50	20	10	0
RNaseP	6/6	6/6	6/6	6/6	6/6	6/6	6/6	6/6
AS1E	6/6	6/6	6/6	5/6	3/6	1/6	0/6	0/6
CU-N2	6/6	6/6	6/6	6/6	1/6	1/6	0/6	0/6
ORF1e	6/6	6/6	5/6	3/6	2/6	1/6	0/6	0/6

Results summary for 20 replicates at 1X LOD

virions/µl	200
RNaseP	20/20
AS1E	20/20
CU-N2	19/20
ORF1e	11/20

Figure 3.6: **SARS-CoV-2 virion limit of detection using RT-LAMP and saliva samples.** **(A)** Saliva samples were spiked with the indicated concentrations of heat-inactivated SARS-CoV-2 virions (top) before being diluted 1:1 with 2X saliva stabilization solution. Samples were then heated at 95^{o}C for 10 minutes and subjected to RT-LAMP at 65^{o}C for 30 minutes in 6 replicates. Each panel represents a unique primer set (listed at the bottom of each panel). The table at the bottom shows a summary of positive reactions observed (yellow). Red box indicates the determined RT-LAMP limit of detection (LOD). **(B)** Saliva samples were spiked with heat-inactivated SARS-CoV-2 virions at a concentration of 200 virions/μL (the limit of detection of our assay) before being diluted 1:1 with 2X saliva stabilization solution. Samples were then heated at 95^{o}C for 10 minutes and 20 replicates of RT-LAMP with the indicated primer sets were incubated at 65^{o}C for 30 minutes. The table at the bottom shows a summary of positive reactions (yellow). Red box indicates our selection of primer sets to advance to subsequent analysis.

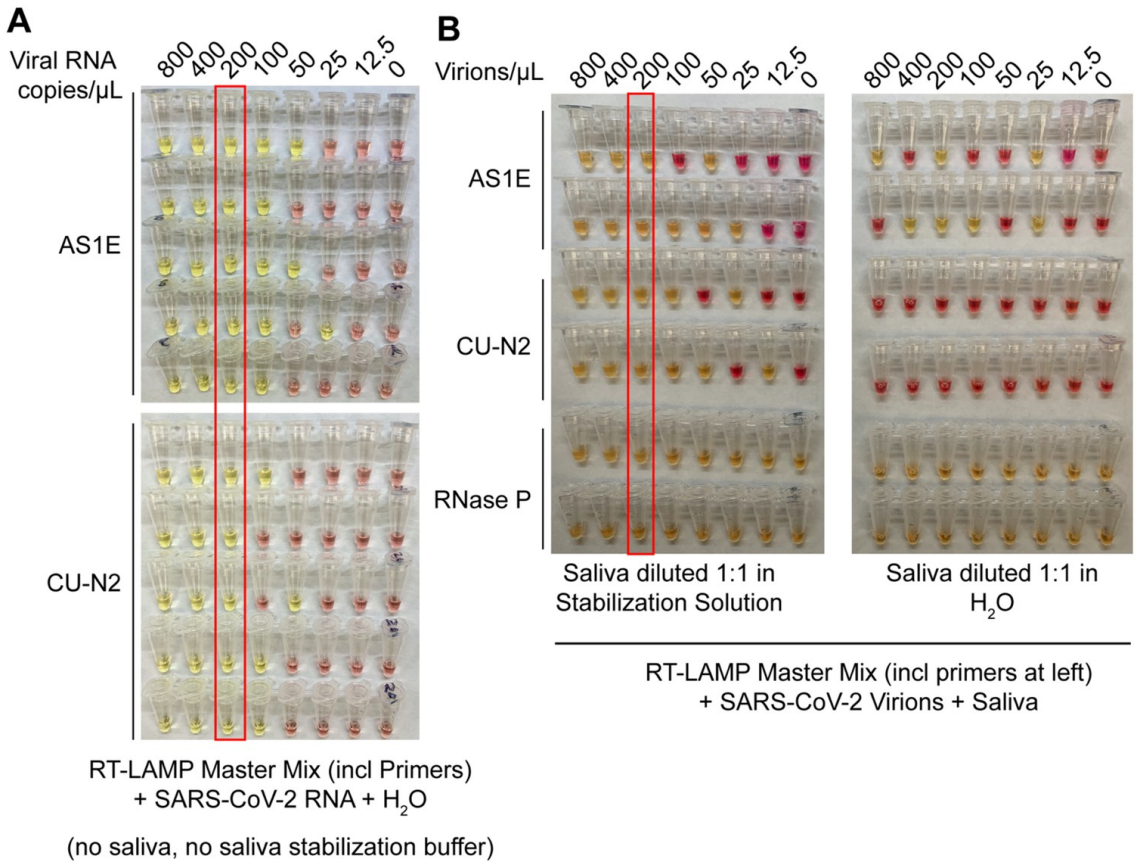


Figure 3.7: **Saliva stabilization solution containing NaOH does not lower sensitivity of colorimetric RT-LAMP detection of SARS-COV-2. (A)** Here, the detection limit of Saliva TwoStep RT-LAMP assay, in the absence of any saliva or saliva stabilization solution, was assessed. This was explored in order to determine whether there might be components of saliva or saliva stabilization solution that inherently lower test sensitivity because they are inhibitory to the RT-LAMP reaction. Here, synthetic SARS-CoV-2 RNA was diluted in nuclease-free water. The diluted RNA was mixed with RT-LAMP reaction mix and incubated at 65ºC for 30 minutes to allow isothermal amplification. Positive reactions turn yellow. Two different primer sets that amplify SARS-CoV-2 were employed, AS1E and CU-N2. The red box indicates the concentration at which positives were identified at least 95% of the time (here, 100% is achieved). That is defined at the limit of detection. Here, it is 200 copies/μL, just as when saliva and saliva stabilization solution is used (see panel B, and data figures in main paper). **(B)** Evaluation of RT-LAMP detection limit in the presence of saliva, but in the presence or absence of saliva stabilization solution. Saliva spiked with heat inactivated SARS-CoV-2 virions at specified concentrations was mixed 1:1 with stabilization solution (left) or nuclease-free water (right) and heated at 95ºC for 10 minutes (RNA liberation) before being incubated at 65ºC for 30 minutes (isothermal amplification). On the left, the saliva stabilization solution achieves a limit of detection of 200 virions /μL. When virions are boiled without the saliva stabilization solution (right), very few reactions turn positive and the pattern is unpredictable, presumably because virions and viral RNA are destroyed.

A. Blinded Study Results Interpretation

AS1E result	CU-N2 result	RNaseP result	Result Interpretation
+	+	+/-	SARS-CoV-2 detected
If only one of the two targets is positive		+/-	Inconclusive result
-	-	+	SARS-CoV-2 not detected
-	-	-	Invalid Result

B. Summary of results from the blinded specimen study

	virions / µL saliva	Number Tested	RNaseP	AS1E	CU-N2	
1X LOD*	200	9	9 (100%)	9 (100%)	9 (100%)	Positive (%)
2X LOD	400	11	11 (100%)	11 (100%)	10 (91%)	
5X LOD	1000	5	5 (100%)	5 (100%)	5 (100%)	
10X LOD	2000	5	5 (100%)	5 (100%)	5 (100%)	
Negative	0	30	30 (100%)	0 (0%)	0 (0%)	
Total		60				

* 1X LOD (limit of detection = 200 virions/µL)

C. Summary of positive and negative agreement

		Contrived Specimen Type		
		Positive	**Negative**	**Total**
Saliva RT-LAMP Test	**Positive**	29	0	**29**
	Inconclusive	1	0	**1**
	Negative	0	30	**30**
	Total	**30**	**30**	**60**
Positive Agreement		97% (29/30)		
Negative Agreement		100% (30/30)		

Figure 3.8: **Blinded sample evaluation.** Plain saliva, or saliva spiked with heat-inactivated SARS-CoV-2 virions at different concentrations, was heated at 95ºC for 10 minutes. Samples were then analyzed using RT-LAMP by a researcher that did not know the true state of each sample. Experiments in figure) For each sample, three reactions were performed as indicated by each triplet of tubes. By looking at the patterns of yellow and pink results in each triplet, samples were scored according to the table below. The true status and observed result of each sample are listed to the right (P = Positive, N = Negative, I = Inconclusive). A white box on the triplet is shown if the sample contained SARS-CoV-2. One sample resulted in inconclusive test result. This sample did have SARS-CoV-2 spiked into it, but one of the SARS-CoV-2 primer sets failed to produce a signal (CU-N2). This failed reaction is still pink (negative) even though the tube has 2xLOD virus. 1X LOD = 200 virions/μL. Summary statistics for this experiment are provided in second and third tables below.

N gene regions of the SARS-2-CoV genome (see methods). From these, all positive samples with viral load above the quantitative RT-PCR detection collected during September 16 – September 25, 2020 (N=278), and size matched 295 negative samples were next re-evaluated with RT-LAMP. Each SARS-CoV-2-positive saliva sample has a Ct value associated with it from the quantitative RT-PCR test conducted by the campus testing team (Figure 3.11). Because positive results in our university screening regimen result in the tested individual being directed to their healthcare provider for confirmatory testing, positively tested individuals were removed from the sampling pool. Thus, most positive samples are from unique individuals, with a few exceptions.

Saliva samples had already been heat inactivated for 30 minutes at 95°C as the initial step of the quantitative RT-PCR protocol. Since the heating component of our Saliva Preparation step had already been performed, an aliquot of the heated saliva sample was transferred into our 2X saliva stabilization solution (without Proteinase K) and then put through the RT-LAMP reaction as described above. For each of the 573 samples, three RT-LAMP reactions were performed with different primer sets: RNaseP (positive control), AS1E, and CU-N2 primer sets (the latter two sets detecting SARS-CoV-2). During this part of the study, we noticed that decreasing the input sample amount (saliva + saliva stabilization solution) from 4 μL to 2 μL in a total reaction volume of 20 μL further increases tolerance of the RT-LAMP reaction color to acidic saliva samples because less saliva is added. We thus reduced the input sample amount to 2 μL when evaluating these human samples. For all 573 samples, RT-LAMP with primers to human RNA positive control (RNaseP) correctly turned positive (yellow).

Specificity: 295 saliva samples that tested negative for SARS-CoV-2 by quantitative RT-PCR were used for evaluation. We re-tested all of those samples with RT-LAMP to evaluate our false-positive rate. Remarkably, for all 295 SARS-CoV-2-negative samples, AS1E and CU-N2 primers sets both universally returned a result of negative, consistent with the quantitative RT-PCR results. Therefore, there was zero false positive, and the test

has a specificity of 100% in this extensive sample set. This shows the strength of our saliva stabilization solution, which mitigates the problem of false-positives in RT-LAMP due to some human saliva samples being naturally acidic.

Sensitivity: We next analyzed 278 SARS-CoV-2-positive saliva samples with viral loads determined based on direct quantitative RT-PCR Ct values using a primer set directed against the nucleocapsid (N) gene of SARS-CoV-2 (see methods). All Ct values reported in this study are from this primer set. We determined the relative viral load of each positive saliva sample based the quantitative RT-PCR standard curve generated by our university testing lab (Figure 3.10). Among all positive samples, 208 (74.8%; AS1E primers) and 182 (65.5%; CU-N2 primers) returned positive RT-LAMP test results (Figure 3.9A). Although both primers sets were still able to detect SARS-CoV-2 RNA below the experimentally determined detection limit (200 virions/μL), we observed a decline in the test sensitivity below such limit (Figure 3.9B). Of the 168 positive samples that contain greater viral load than RT-LAMP limit of detection, 158 (94%; AS1E primers) or 142 (85%; CU-N2 primers) returned positive RT-LAMP test results (Figure 3.9A). In Figure 3.9C, we summarize the performance of each primer set in both this test of human saliva samples and in the spiked in virion experiments described above (Figure 3.8). The observed limit of detection of the AS1E primer set was determined from this data to be 266 virions/microliter. The strong congruence with our prior estimate of 200 virions/microliter demonstrates that heating for 30 minutes prior to adding stabilization solution and using 2 µL of saliva plus stabilization solution, instead of 4 µL, both have very little effect. Because the AS1E primer set performs best throughout our study, we include that as the main primer set in our final test configuration, which we refer to as the Saliva TwoStep test. However, the CU-N2 primer set still performs well and can be used when it is desirable to detect a second region of the SARS-CoV-2 genome.

Test sensitivity as a function of viral load in the sample: For both primer sets, we calculated the sensitivity (positive agreement with quantitative RT-PCR) and specificity

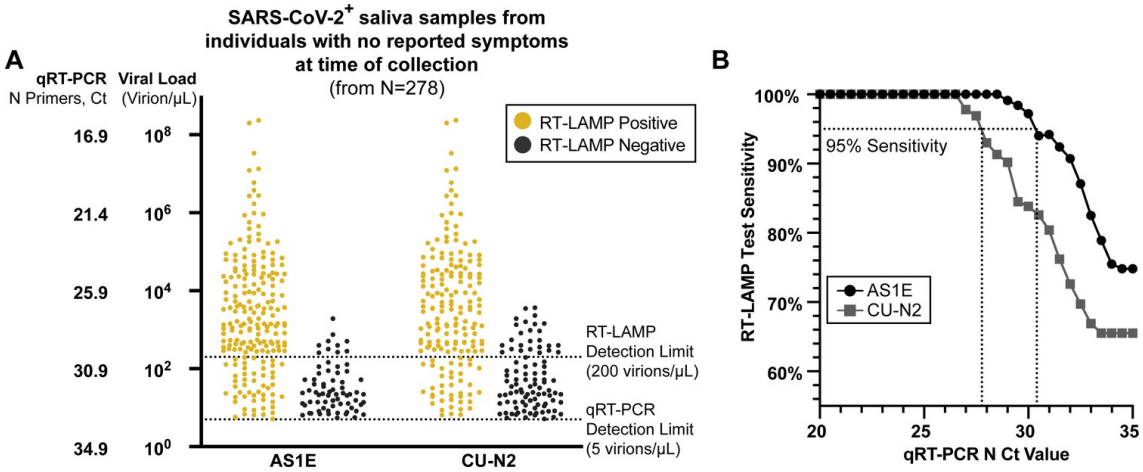
Table 3.1: **Summary of RT-LAMP evaluation in human samples.**

			RT-LAMP				
				AS1E		CU-N2	
			No. of Samples	No. of Positives	Agreement	No. of Positives	Agreement
Quantitative RT-PCR (SARS-CoV-2 N)	Negative		295	0	295/295 (100%)	0	295/295 (100%)
	Positive (Levels of Viral Load: Virions/μL)	4000	82	82	82/82 (100%)	82	82/82 (100%)
		2000	97	97	97/97 (100%)	94	94/97 (96.9%)
		1000	118	117	117/118 (99.2%)	110	110/118 (93.2%)
		800	123	122	122/123 (99.2%)	112	112/123 (91.1%)
		400	143	139	139/143 (97.2%)	129	145/173 (90.2%)
		200	168	158	158/168 (94.0%)	142	142/168 (84.5%)

(negative agreement with quantitative RT-PCR) of the RT-LAMP test at various levels of viral load cutoffs (Figure 3.9B, Table 3.1). The differences in the observed limit of detection between the two SARS-CoV-2 primer sets could reflect the differences in the primer efficiencies, as well as the dynamics in relative viral transcript abundance [116].

3.4.7 Assessment of Saliva TwoStep against a nasal swab test.

Of the 278 SARS-CoV-2-positive saliva samples analyzed above, 54 also had matched nasal samples collected no more than two days later. In some cases, individuals may have developed symptoms by the time follow-up nasal swabs were taken, so we can make no claims about symptomatic status at the time of nasal swab. We next compared the results of the Saliva TwoStep test with the result obtained by the Quidel Lyra direct nasal swab RT-PCR test. Compared to the quantitative RT-PCR on saliva results, the RT-LAMP produced three false negative in this sample set, whereas the Lyra nasal swab test produced eight (Figure 3.12A). However, this is still remarkably consistent given that this comparison involves three degrees of freedom: biosample (saliva versus nasal swab), test modality (RT-PCR versus RT-LAMP), and days between saliva and nasal samples collection (up to two days apart). A summary of how these first two degrees of freedom affect test congruency are shown in Figure 3.12B.



C

Estimates of Saliva TwoStep sensitivity and specificity summarized from this study

	AS1E Primer Set		CU-N2 Primer Set	
	Sensitivity	Specificity	Sensitivity	Specificity
Blinded virion spike-in samples	100% out of 30 SARS⁺ samples	100% out of 30 SARS⁻ samples	97% out of 30 SARS⁺ samples	100% out of 30 SARS⁻ samples
Saliva from individuals with no reported symptoms at sample collection	94% out of 168 SARS⁺ samples (> 200 virions/µL)	100% out of 295 SARS⁻ samples	85% out of 168 SARS⁺ samples (> 200 virions/µL)	100% out of 295 SARS⁻ samples
	75% out of 278 SARS⁺ samples (> 5 virions/µL)		65% out of 278 SARS⁺ samples (> 5 virions/µL)	

Figure 3.9: **Evaluation of RT-LAMP on SARS-CoV-2-positive saliva samples from individuals with no reported symptoms at the time of sample collection.** **(A)** We re-analyzed university saliva samples that had been previously analyzed for SARS-CoV-2 using quantitative RT-PCR with a primer set against the N gene of SARS-CoV-2 (see methods). The remaining saliva was mixed 1:1 with 2X saliva stabilization solution (without Proteinase K) and re-tested using RT-LAMP. The results of RT-LAMP are compared to relative saliva viral load determined by quantitative RT-PCR. The figure shows the distribution of the viral load of all 278 positive saliva samples separated by the corresponding RT-LAMP reaction results with either the AS1E or CU-N2 primer-set. **(B)** Saliva TwoStep RT-LAMP test sensitivity as a function of the cycle threshold (Ct) from the quantitative RT-PCR results of the corresponding SARS-CoV-2 positive saliva samples. **(C)** A summary of the sensitivity and specificity of the Saliva TwoStep test from the blinded sample evaluation described above and shown in Figure 3.11 (top), and from both the data in panel A (bottom).

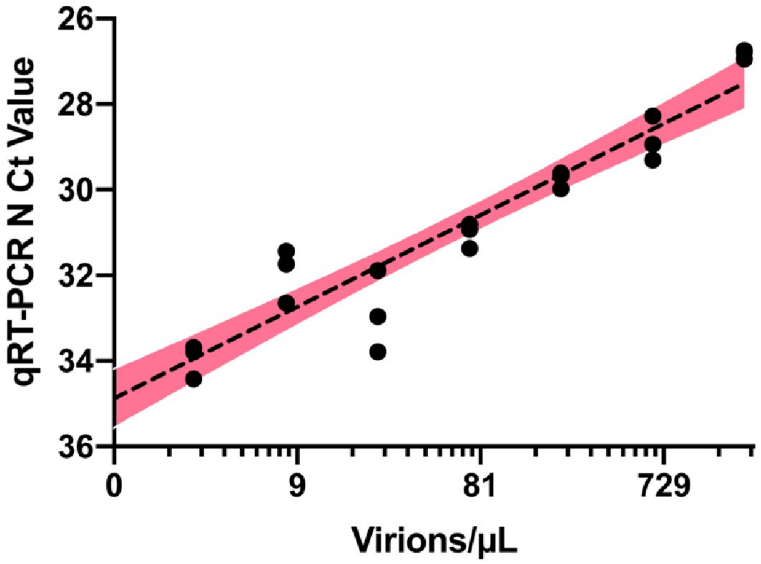


Figure 3.10: **Quantitative RT-PCR standard curve used to determine the Ct value to virion/μL calculation** 10,000 copies/µL of heat deactivated SARS-CoV-2 virus was spiked into negative saliva specimens from 6 different volunteers and incubated for 30 minutes at 95ºC. Samples were diluted to indicated concentrations using heat-treated saliva without SARS-CoV-2 addition from the same individual. The standard curve for the primer set targeting SARS-CoV-2 N gene is generated from the linear regression analysis and is illustrated with 95% confidence interval.

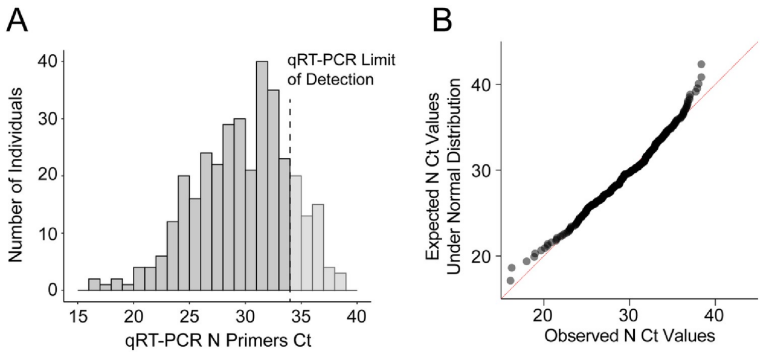


Figure 3.11: **Near normal distribution of quantitative RT-PCR raw Ct values of SARS-CoV-2 N gene from positive individuals.** Between September 16 – September 25, 2020, 8836 saliva samples were screened for SARS-CoV-2 using the direct quantitative RT-PCR method. **(A)** The amplification of SARS-CoV-2 N gene is detected in 347 samples and the Ct value distribution is illustrated. Samples with Ct values above the qRT-PCR limit of detection (Ct¡34, Dark Grey, N=278) were considered SARS-CoV-2 positive and used for RT-LAMP TwoStep validation. **(B)** Quantile-quantile plot of the SARS-CoV-2 N Ct values indicates near normal distribution within the linear range of qRT-PCR (D'Agostino test, K2=9.07, p-value=0.011).

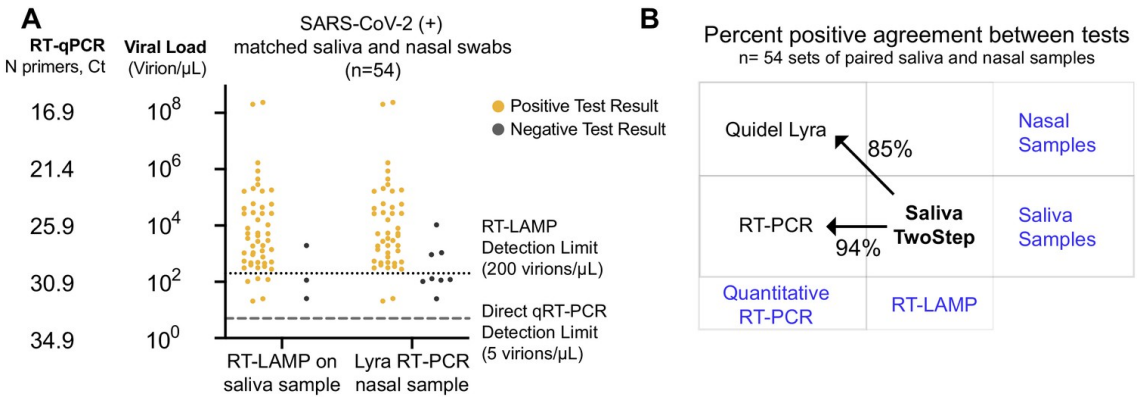


Figure 3.12: **Assessment of Saliva TwoStep against a nasal swab test. (A)** Matched nasal swabs and saliva from 54 individuals were analyzed (all of whom were SARS-CoV-2 positive at the time that these samples were collected, as verified by saliva direct quantitative RT-PCR test). Nasal swab samples from the same individuals were collected within 2 days of positive saliva test, and tested using the Quidel Direct Lyra RT-PCR test. The saliva samples from those same individuals were tested with the Saliva TwoStep test. Data points represent individuals (n = 54), and the corresponding test result is color-coded: positive, yellow; negative, grey. **(B)** Positive test agreement between Saliva TwoStep and the two comparator tests. The nature of the sample used by each test (nasal swab or saliva), and the test chemistry (quantitative RT-PCR or RT-LAMP) are delineated.

3.4.8 Final test conditions

From the experiments described above we selected the final optimized conditions for our Saliva TwoStep test. The two steps have an end-to-end processing and analysis time of approximately 45 minutes (Figure 3.13). For additional application details regarding the testing station setup, sample collection, and overall workflow of employing this test for community screening, please refer to the appendix below.

Step 1. Prepare Saliva: Collect saliva, combine 1:1 with 2X saliva stabilization solution and incubate at 95^{o}C for 10 minutes. Note: We have determined that performing a heating step at 95^{o}C for 30 minutes in a water bath, before addition of the saliva stabilization solution, also works reasonably well. However, in this case Proteinase K must be omitted.

Step 2. Detect Virus: Incubate at 65^{o}C for 30 minutes: 2 μL diluted saliva from step 1, 10 μL 2X NEB Colorimetric RT-LAMP enzyme mix, 6 μL of nuclease-free water and 2 μL 10x primer mix for a final reaction volume of 20 μL.

Step 3. Reaction Inactivation (optional): Stop reaction at 80^{o}C for 2 minutes. This stabilizes color so that results can be analyzed at a later time. The multiple heating steps here may be programmed into a thermal cycler for maximum convenience, but this is not necessary.

3.5 Discussion

There are several advantages to the SARS-CoV-2 Saliva TwoStep RT-LAMP screening approach described herein: 1) The use of saliva eliminates invasive nasal swab-based sampling, which requires special supplies and causes discomfort. 2) We optimized saliva stabilization solution that allows for the neutralization of a broad range of naturally acidic saliva samples while maintaining compatibility with a colorimetric RT-LAMP assay. The solution also helps preserve saliva samples for at least four days before processing and lowers saliva viscosity. 3) We determined the optimal sample heating condition that liberates the

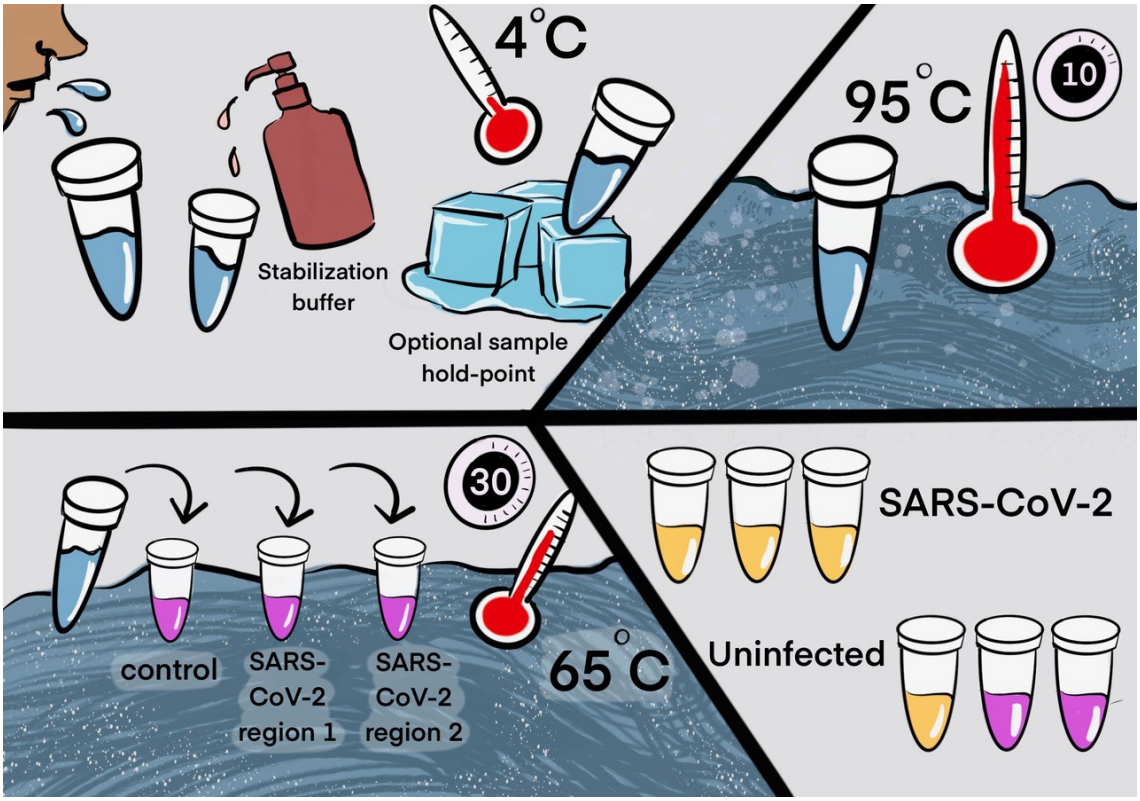


Figure 3.13: **Two step detection of SARS-CoV-2 in saliva. Upper half) Step 1**: Prepare Saliva. Person provides 1 mL of saliva, and 1 mL of 2X saliva stabilization solution is then added to it. (This sample can be processed immediately or stored in the refrigerator at 4^{o}C for at least 4 days.) The mixture is heated at 95^{o}C for 10 minutes. This step serves to neutralize the pH of saliva, liberate viral RNA from virions in the saliva, and inactivate virions for safe handling (although appropriate safety precautions should always be taken). We have determined that performing a heating step at 95^{o}C for 30 minutes in a water bath before addition of the saliva stabilization solution also works equally well. However, in this case Proteinase K must be left out of that solution.
Lower half) Step 2: Detect Virus. 2 μL of stabilized saliva from step 1 is pipetted into each of three test tubes pre-filled with the RT-LAMP master mix and primers. The only thing different between the three tubes is the primer set included, with each set targeting either the human positive control RNA or a region of SARS-CoV-2 RNA, as indicated. After incubation, the reaction will turn from pink to yellow if the target RNA is present in saliva. An example of a positive and a negative test are shown. Graphic by Annika Rollock.

host and viral RNA with minimal degradation. The simple heating step increases biosafety and avoids formal RNA extraction procedures. 4) For RT-LAMP, we incorporated additional primers based on up-do-date SARS-CoV-2 genome databases and identified primers allowing efficient target amplification. These primers are expected to work on most of all viral variants currently circulating (Figure 3.5). Overall, with the simplified two steps of saliva preparation and virus detection, the test has a rapid sample-to-result turnaround time of 45 minutes.

Through the optimization process, we identified other potential sources of false positive results and provided a detailed summary for troubleshooting (Methods and Appendix). In addition, from our experience of the actual deployment of this screening test, we summarized the standard operational procedures for saliva sample collection, including the design of a stabilization solution dispensing apparatus to preserve samples while avoiding environmental contamination and protecting workers (Figure 3.14; Appendix). By strictly following these application notes, we completely avoided false positive results during the evaluation of a large cohort of human saliva samples, achieving 100% specificity. We also evaluated the RT-LAMP test performance based on the experimentally determined limit of detection. Using SARS-CoV-2 positive human saliva samples, we confirmed that the RT-LAMP test can consistently identify infected individuals with 94% sensitivity.

During the test development and optimization, we have also explored additional methods that may help enhance the RT-LAMP test performance and consistency. Previous work suggests that the addition of 40 mM of guanidine chloride in the RT-LAMP reaction mix could increase RT-LAMP amplification efficiency [124]. However, we did not observe similar enhancement when included in our experiments. To further prevent carry over contamination [152], the usage of dUTP and uracil-DNA-glycosylase-containing RT-LAMP reaction mix can be considered. Previous studies have shown that the addition of this alternative master mix does not affect the RT-LAMP limit of detection [134, 153].

Saliva TwoStep requires less sample processing, reaction incubation time, and labora-

tory overhead as compared to quantitative RT-PCR methods. The result is the ability to run significantly more tests with a given amount of resources. Based on these observations, we conclude that the Saliva TwoStep test described herein can be used as a SARS-CoV-2 screening tool to reliably identify highly infective individuals with minimal laboratory setup, potentially serving as a tool for effective SARS-CoV-2 surveillance at the community level. This RT-LAMP testing offers many solutions to a nation-wide shortage of COVID-19 testing. With minimal set-up this test could be performed in diverse settings such as factories, office buildings, or schools.

3.6 Methods

3.6.1 RT-LAMP primer design and preparation

Regions of the SARS-CoV-2 genome that are conserved among strains were identified using genome diversity data from NextStrain (nextstrain.org/ncov/global). Next, nucleotide-BLAST (blast.ncbi.nlm.nih.gov) was used to filter out genome sequences that share high sequence homology with other seasonal coronavirus genomes. Finally, PrimerExplorer V5 (primerexplorer.jp/e/) was used to design RT-LAMP primers targeting the specific regions of SARS-CoV-2 genomes. The F3, B3, FIP, BIP, Loop F and Loop B primers were selected for optimal melting temperature and complementarity using A plasmid editor (ApE). All primers were ordered from IDT in desalted form. In all cases, a 10X concentration of primer sets was made containing 16 μM FIP and BIP primers, 4 μM LF and LB primers, and 2 μM F3 and B3 primers. All primers should be ordered with HPLC purification, which ensures the yield and avoids cross contamination from other SARS-CoV-2-related synthesis projects being run on the same equipment at the oligo synthesis facilities (which can lead to false positives). This is particularly a problem during a pandemic where these facilities are handling many oligo synthesis orders focused on the same pathogen[154]. It is also recommended that you communicate with the primer synthesis company to inform them that primers are

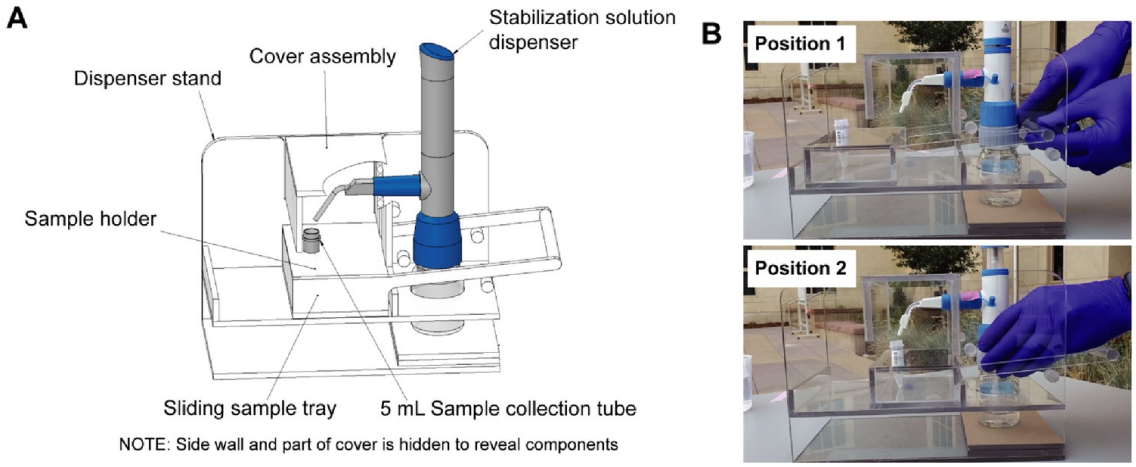


Figure 3.14: **Diagrams of the saliva stabilization solution dispensing apparatus. (A)**: CAD model of dispensing apparatus showing components. Custom solution dispensing apparatus fabricated from machined and solvent welded .236in polycarbonate (Tuffak). Polycarbonate is chosen for visibility, strength, and ability to withstand cleaning solvents such as ethanol. This device prevents the need for staff to directly handle uncapped and potentially infectious sample prior to inactivation, limits splash and aerosol exposure risk, and prevents cross-contamination of samples during solution addition step. **(B)** Diagram illustrates the operation of the dispensing apparatus. Position 1: Tray is extended towards the testing participant and sample tube is seated in tray. Bottle containing stabilization solution and assembled with bottle top dispenser is seated in back section of apparatus. Staff moves tray towards themselves by gently pulling on handle until the tray is seated against back wall of the apparatus. Position 2: Sample tray is positioned against the back wall of the apparatus. This brings the sample tube and dispenser nozzle into a set orientation underneath the removable cover assembly. Sample collector uses bottle top dispenser to add 1mL of stabilization solution to sample then pushes sample tray back to the testing participant.

intended for use with a SARS-CoV-2 screening test. Several companies have dedicated facilities for minimizing cross-contamination of SARS-CoV-2 templates. In addition, primers should be diluted in nuclease-free water, instead of Tris-EDTA (TE) buffer, which will also inhibit pH change that takes place during RT-LAMP.

Primer Set Name	LAMP Primer Component	Primer Sequence (5' - 3')
"RNaseP" amplifies human RNA for positive control	F3	TTGATGAGCTGGAGCCA
	B3	CACCCTCAATGCAGAGTC
	Loop F	ATGTGGATGGCTGAGTTGTT
	Loop B	CATGCTGAGTACTGGACCTC
	FIP	GTGTGACCCTGAAGACTCGGTTTTAGCCACTGACTCGGATC
	BIP	CCTCCGTGATATGGCTCTTCGTTTTTTTCTTACATGGCTCTGGTC
"AS1E"	F3	CGGTGGACAAATTGTCAC
	B3	CTTCTCTGGATTTAACACACTT
	Loop F	TTACAAGCTTAAAGAATGTCTGAACACT
	Loop B	TTGAATTTAGGTGAAACATTTGTCACG
	FIP	TCAGCACACAAAGCCAAAAATTTATTTTTCTGTGCAAAGGAAATTAAGGAG
	BIP	TATTGGTGGAGCTAAACTTAAAGCCTTTTCTGTACAATCCCTTTGAGTG
"CU-N2" developed herein	F3	CGGCAGTCAAGCCTCTTC
	B3	TTGCTCTCAAGCTGGTTCAA
	Loop F	*This set does not require a Loop F primer*
	Loop B	ATGGCGGTGATGCTGCTCTT
	FIP	TCCCCTACTGCTGCCTGGAGCGTTCCTCATCACGTAGTCG
	BIP	TCTCCTGCTAGAATGGCTGGCATCTGTCAAGCAGCAGCAAAG
"ORF1e" developed herein	F3	GGCTAACTAACATCTTTGGC
	B3	GTCAGCACACAAAGCCAA
	Loop F	TCTTCAAGCCAATCAAGGAC
	Loop B	TTGTCGGTGGACAAATTGT
	FIP	TCTCTAAGAAACTCTACACCTTCCTTTTTACTGTTTATGAAAAACTCAAACC
	BIP	TATCTCAACCTGTGCTTGTGAAATTTTAGAATGTCTGAACACTCTCCT

3.6.2 SARS-CoV-2 RNA and virion standards

Synthetic SARS-CoV-2 RNA control (Twist Bioscience #102019) was obtained and its copy number of $1x10^6$ copies/μL was confirmed using quantitative RT-PCR in conjunction with a DNA plasmid control containing a region of the N gene from the SARS-CoV-2 genome (IDT #10006625). Heat-inactivated SARS-CoV-2 virion control (ATCC #VR-1986HK) was obtained and its concentration of $3.75x10^5$ virions/μL was confirmed using quantitative RT-PCR in conjunction with both the synthetic SARS-CoV-2 RNA control and a DNA plasmid control containing a region of the N gene from the SARS-CoV-2 genome. SARS-CoV-2 RNA was added to saliva samples after being mixed 1:1 with saliva stabilization solution and heated at 95^{o}C for 10 minutes, unless stated otherwise, whereas heat-inactivated SARS-CoV-2 virions were added to saliva samples and mixed 1:1 with saliva stabilization solution

before being heated. Concentrations reported throughout this study represent the final concentration of standards in saliva before it was mixed 1:1 with 2X saliva stabilization solution.

3.6.3 Saliva preparation with heat and stabilization solution

When making the 2X saliva stabilization solution, we offer several key pointers: 1) Use TCEP-HCl (GoldBio #TCEP10). The -HCL form must be used to produce the correct final stock pH. 2) Use EDTA, 0.5 M, pH 8.0 (Sigma-Aldrich #324506). It is important to use a pH 8.0 stock solution, otherwise this also affect the pH of the final stabilization solution. 3) Use lyophilized Proteinase K (Roche #3115879001). It is important to use the lyophilized form. Liquid forms will contain Tris, which inhibits the pH change during the RT-LAMP reaction. 4) 10 M NaOH was prepared by dissolving NaOH pellets (Sigma-Aldrich #221465) into nuclease-free water, before being added to the 2X solution to reach the correct concentration. The following is the exact recipe that we used to create a 100mL stock of 2X saliva stabilization solution:

Components	Amount Mixed	Final Concentration in 2X stock solution
TCEP-HCl (GoldBio #TCEP10)	143.3 mg	5 mM
0.5 M EDTA, pH 8.0 (Sigma-Aldrich #324506)	400 µL	2 mM
10 M NaOH (Sigma-Aldrich #221465)	290 µL	29 mM
Proteinase K (Roche #3115879001)	10 mg	100 µg/mL
Nuclease-free water	To 100 mL	

Saliva samples (1 mL) were collected in sterile, nuclease-free 5 mL conical screw-cap tubes (TLD Five-O #TLDC2540). 2X saliva stabilization solution described above was then added at a 1:1 ratio. Samples were shaken vigorously for 5 -10 seconds and incubated at 95^{o}C for 10 minutes. Samples were then placed on ice before being used in downstream

analyses (Detailed sample collection procedure is described in the Appendix).

3.6.4 Real-time RT-LAMP

For each reaction, 10 μL WarmStart LAMP 2X Master Mix (NEB #E1700) was combined with 1 μL 20X EvaGreen Dye (Biotium #31000), 2 μL 10X primer mix, and 3 μL DEPC-treated water. The combined reaction mix was added to MicroAmp Optical 96-Well Reaction Plate (ThermoFisher #N8010560) and then 4 μL processed saliva sample was added. The reaction was mixed using a multi-channel pipette and incubated in Applied Biosystems QuantStudio3 Real-time PCR system. The reaction proceeded at 65^{o}C for 30 minutes with fluorescent signal being captured every 30 seconds. The results were visualized and analyzed using ThermoFisher's Design and Analysis software.

3.6.5 Colorimetric RT-LAMP

WarmStart Colorimetric LAMP 2X Master Mix (NEB #M1800) was used in all colorimetric RT-LAMP reactions. Each reaction was carried out in a total of 20 μL (10 μL WarmStart Master Mix, 2 μL 10X primer mix, 4 μL processed saliva sample, and 4 μL DEPC-treated water). Reactions were set up in PCR strip tubes on ice. Saliva template was added last and tubes were inverted several times to mix samples and briefly spun down in a microfuge. Reactions were incubated in a thermal cycler at 65^{o}C for 30 minutes and then deactivated at 80^{o}C for 2 minutes. The incubation was carried out without the heated lid to simulate a less complex heating device. Images of reactions were taken using a smartphone. For the community deployment of this assay, 2 μL of processed saliva was used instead of 4 μL.

3.6.6 Testing of University Samples

The University of Colorado Boulder SARS-CoV-2 screening test was loosely based on a published quantitative RT-PCR reaction performed directly on saliva [94], which has a

limit of detection of 5 virions/μL. Some modifications were made, as described here. For sample collection, individuals were asked to collect no less than 0.5 mL of saliva in a 5 mL screw-top collection tube. Saliva samples were heated at 95^{o}C for 30 minutes to inactivate the viral particles for safe handling, and then placed on ice or at 4^{o}C. For quantitative RT-PCR analysis, the university testing team transferred 75 μL of saliva into a 96-well plate, where each well had been pre-loaded with 75 μL 2xTBE buffer supplemented with 1% Tween 20. (The remaining saliva in the 5 mL collection tube proceeded to RT-LAMP testing as described in the next paragraph). Next, 5 μL of this diluted sample was added to a separate 96-well plate containing 15 μL reaction mix composed of: TaqPath 1-step Multiplex Master Mix (Thermo Fisher A28523), nuclease-free water, and triplex primer mix consisting of primer and probe sets targeting SARS-CoV-2 E and N genes and human RNase P gene (sequence and concentration specified in the table below). The reactions were mixed, spun down, and loaded onto a Bio-Rad CFX96 or CFX384 qPCR machine. Quantitative RT-PCR was run using the standard mode, consisting of a hold stage (25^{o}C for 2 minutes, 50^{o}C for 15 minutes, and 95^{o}C for 2 minutes) followed by 44 cycles of a PCR stage (95^{o}C for 3 seconds, 55^{o}C for 30 seconds, with a 1.6^{o}C/sec ramp up and ramp down rate). Only Ct values from the N primer set are reported in the study herein, and used to calculate relative sample viral load based on the standard curve shown in Figure 3.10.

Leftover samples from this testing procedure were then tested with RT-LAMP. 50 μL of saliva samples was transferred and mixed into a 96-well plate, pre-loaded with 50 μL 2X saliva stabilization solution without proteinase K (5 mM TCEP-HCL, 2 mM EDTA, 29 mM NaOH, diluted in DEPC-treated water). 2 μL of diluted saliva samples were transferred into 8-strip PCR tubes containing RT-LAMP reaction mixture (enzymes and primers). For each sample, three RT-LAMP reactions were carried out to amplify human RNaseP as a control and AS1E and CU-N2 for SARS-CoV-2. The reactions were incubated at 65^{o}C for 30 minutes followed by inactivation at 80^{o}C for 2 minutes on a thermal cycler (Bio-RAD T100). A color change from pink to yellow was observed visually to interpret results.

TaqMan Primer/ Probe Set Target	Primer or Probe Name	1X Concentration	Sequence (5' - 3')
SARS-CoV-2 E gene	E_Sarbeco_F1 (IDT 10006888)	400 nM	ACAGGTACGTTAATAGTTAATAGCGT
	E_Sarbeco_R2 (IDT 10006890)	400 nM	ATATTGCAGCAGTACGCACACA
	E_Sarbeco_P (IDT Custom)	200 nM	TexRd-ACACTAGCCATCCTTACTGCGCTTCG-IAbRQSp
SARS-CoV-2 N gene	nCOV_N1_F (IDT 10006830)	500 nM	GACCCCAAAATCAGCGAAAT
	nCOV_N1_R (IDT 10006831)	500 nM	TCTGGTTACTGCCAGTTGAATCTG
	nCOV_N1_P (IDT Custom)	250 nM	HEX-ACCCCGCAT-ZEN-TACGTTTGGTGGACC-IABkFQ
Human RNase P	RNaseP_F (IDT 10006836)	50 nM	AGATTTGGACCTGCGAGCG
	RNaseP_R (IDT 10006837)	50 nM	GAGCGGCTGTCTCCACAA GT
	RNase_P_P (IDT 10006838)	50 nM	FAM-TTCTGACCT-ZEN-GAAGGCTCTGCGCG-IABkFQ

3.7 Appendix: Saliva TwoStep SARS-Cov-2 Screening Test Implementation Notes

3.7.1 Recommended procedures for sample collection

When deploying Saliva TwoStep RT-LAMP screening test, it is important to set up saliva sample collection sites that allow large numbers of participants to move through the sample collection process quickly and smoothly. In addition, it is important to exercise extra precautions to avoid sample cross contamination as well as the exposure of the sample or stabilization solution. With this in mind, we have designed and optimized a saliva sample collection workflow that utilizes a customized stabilization dispensing apparatus (Figure 3.14):

1. At the designated sample collection site, the screening test participants retrieve a 5 mL screw cap tube (MTC Bio #C2530) and collect passive drool into the 5mL tube until liquid saliva reaches the 1mL graduation mark. Bubbles do not contribute towards the 1mL volume.

2. After collecting saliva, the testing participants submerge capped 5 mL collection tube in a 250mL beaker with 70% ethanol to decontaminate the surface. They then remove

the tube, carefully uncap it, and place it in the slot of the dispensing apparatus sample tray while in Position 1 (Figure 3.14).

3. Staff then move the sample tray towards themselves by gently pulling on sample tray handle until it is seated against the back wall and the sample tube is underneath the cover assembly and centered under the dispenser nozzle (Figure 3.14B Position 2). This partially enclosed space limits the potential risk of splashes and aerosols during this solution addition step.

4. Staff then use the bottle top dispenser (Fisher Scientific #13681527) to gently add 1mL of stabilization solution into sample tube (for an approximate 1:1 ratio of sample to solution). The apparatus holds the dispenser nozzle and sample tube in a fixed orientation to prevent cross contamination during this step.

5. Staff slide the sample tray containing the collection tube back towards the participant. The participant re-approaches to cap their sample with the screw-top lid, shakes it vigorously for 5-10 seconds to mix, cleans the surface with a wipe or by dunking in disinfectant, and places it on ice.

6. Staff then heat-inactivates the sample on-site by incubating it in a 95°C water bath or heat block (heat block preferred for minimizing spill risk) for 10 minutes.

7. Before the next testing participant approaches, staff sprays the sample tray with disinfectant.

8. Staff subsequently stores the inactivated sample on ice in a cooler and then transports it to Saliva TwoStep RT-LAMP testing area (see below).

3.7.2 Biosafety note

Staff involved in saliva collection should wear all appropriate PPE including a fit-tested N95 mask. Regular surface decontamination is performed with 70% ethanol or bleach in the case of spills. Heating elements and cords are secured and situated away from foot traffic. Collection takes place outdoors for ventilation purposes whenever possible. Subjects

maintain 10-foot distance from each other while unmasked and producing samples. Hand sanitizer is provided before and after collection procedure. Protocols were approved by the Institutional Biosafety Committee.

3.7.3 Collection site material list

Staff set up a table with the following supplies: (1) 250mL plastic beaker with 70% ethanol, (1) 250mL plastic beaker with 10% bleach, (2) spray bottles with 70% ethanol, (1) dispenser apparatus (custom polycarbonate device) with bottle top reagent dispenser (Fisher) and 100mL glass bottle with stabilization solution, (1) ice bucket with ice, (1) water bath with tube rack and temperature probe, (1) digital timer, (1) cooler with ice. Additional items include paper towels, trash receptacles, spill kits, power cords, hand sanitizer, and additional PPE, ethanol, and bleach. Quantities can be scaled up as needed.

3.7.4 Recommended testing lab setup

3.7.4.1 Controlling acidic and variable human saliva samples

We found that the biggest obstacle to implementing the colorimetric RT-LAMP assay is the variability in the reaction pH condition. First, false-positive signal can result from saliva samples that are naturally acidic. We spent significant time addressing this issue and ultimately found that all samples must be rendered basic as described in the article in order to set the correct threshold for specificity in the test. This was achieved through a titration series of sodium hydroxide in the saliva stabilization solution to find the optimal concentration that would ensure all RT-LAMP reactions start pink and are still capable of turning yellow if amplification occurs. However, note that other components in the saliva stabilization solution (EDTA, pH=8.0, and TCEP-HCl) have also played a role in establishing the optimal pH. Additional saliva stabilization solution optimization might be needed if other forms of these components are used. Second, the colorimetric RT-LAMP reaction relies

on phenol red to detect the pH change during the amplification. Any additional buffering agents, such as tris-acetate or tris-borate that are commonly present in laboratory reagents, should be avoided to prevent potential false negative signals, as these buffering agents tend to inhibit the pH change.

3.7.4.2 Controlling reaction acidification by carbonic acid

See also Thi et al [155] for additional, excellent advice.

A second issue that has to be carefully controlled is the exposure of reaction components to the surrounding environment. When carbon dioxide from the atmosphere dissolves in water, it creates carbonic acid, which if present in high enough quantities can trigger the phenol red to turn yellow regardless of reaction state. Control measures should be implemented in three ways: First, we advise preparing the reaction mix (RT-LAMP master mix, primers, and water) right before sample loading. This is to prevent background amplification as well as the acidification of the reaction mix due to exposure to air. For this reason, we advise against the use of 96-well plates. Additionally, dry ice should be avoided or completely isolated from the reaction components during sample transportation, as the exposure to the excessive carbon dioxide could also lead to acidification of the reaction mix. Second, during the 30-minute 65^{o}C incubation, it is essential to completely seal off the reaction tubes to prevent vaporization of the reaction mix, as well as the infiltration of the water vapor if a water bath is used. We have noticed that an incomplete seal could lead to false positive signals. Last, because RT-LAMP amplification is highly robust, the test is very sensitive to contamination [152]. Therefore, opening of reaction tubes after RT-LAMP has occurred should be strictly avoided as these tubes contain a large amount of target DNA. Alternatively, the NEB WarmStart LAMP 2X Master Mix with UDG (NEB M1804) can be used to eliminate DNA contamination. Through these results are not shown, we have verified the same limit of detection can be reach using this alternative master mix.

3.7.4.3 Controlling laboratory-based contamination

See also Davidi et al [152] for additional, excellent troubleshooting advice. When carrying out the RT-LAMP SARS-CoV-2 screening test at scale, it is critical to assign isolated workstations, each containing their own set of laboratory equipment such as pipettes, centrifuges, vortexes, and cleaning supplies. This equipment should never move between stations and be regularly decontaminated using a detergent based cleaning solution such as 10% bleach or any other commercially available solution designed to eliminated nucleic acid contaminants. Additionally, special care should be taken by laboratory staff to regularly replace gloves if moving backwards from the following workstations:

Workstation 1: Setting up master mixes. This workstation is dedicated to making and aliquoting the master mix containing the RT-LAMP enzymes, primers, and water. All reagents should be centrifuged and spun down after thawing. As mentioned above, master mix should be made and aliquoted shortly before addition of saliva (Workstation 2) to avoid carbon dioxide solubilization due to atmospheric exposure.

Workstation 2: Adding saliva samples to aliquoted RT-LAMP reaction tubes. This workstation is dedicated to handling the processed saliva samples. Once saliva samples are added to the aliquoted RT-LAMP reaction tubes, care should be taken to ensure that an appropriate seal is established (e.g. dome cap strips) to minimize airflow during reaction incubation. This workstation should include two micropipettes capable of pipetting a volume of 2 µl. One pipette can be used for saliva samples, while the other pipette should be used exclusively for pipetting any RNA controls (e.g. *in vitro* transcribed SARS-CoV-2 RNA).

Workstation 3: RT-LAMP reaction incubation and results reporting. This workstation contains the heating element (e.g. heat block, thermal cycler) where RT-LAMP reactions are incubated. This Workstation has the highest risk of contamination, since the RT-LAMP reaction products will be in high abundance and can themselves serve as a template in subsequent reactions. When reactions are removed from the heating element, they should

immediately be analyzed, results recorded, and then reaction tubes should be disposed in a container with a lid. Never carry completed reaction tubes to any other part of the lab. Never open completed reaction tubes for any reason. Any laboratory technician that has entered Workstation 3 should dispose of their gloves before returning to any other part of the lab.

Chapter 4

A universal immune response to diverse pathogenic infections can be measured in human saliva

---

Adapted from: Yang Q, Meyerson NR, Paige CL, Morrison J, Clark SK, Fattor WT, Decker C, Steiner H, Lian E, Perera R, Poeschla E, Parker R, Dowell RD, Sawyer SL. A universal immune response to diverse pathogenic infections can be measured in human saliva. In preparation. 2022.

4.1 Contributions

This is a highly collaborative project among several laboratories across University of Colorado Boulder and Anschutz Medical Campuses. I joined this project at its early days, helped layout the logistics of sample collection and processing, and was responsible for some of the RNA purifications from *in vitro* and clinical samples. Further, I carried out all of the computational analyses of sequencing data, helped design the multiplex TaqMan RT-qPCR panel, and performed the subsequent RT-qPCR analyses. With the help of my mentors, I generated the figures and wrote the following manuscript. But this project would not be possible without all of the wonderful people listed here as the co-authors. I am grateful for the help from Dr. James Morrison (Poeschla Lab), Camille Paige, Will Fattor and Kyle Clarks (Sawyer Lab) for collecting the valuable clinical and student samples and their involvement in the initial study design. Dr. Carolyn Decker and Halley Steiner (Parker Lab) were the

key drivers for figuring out how to process the saliva samples for RNA sequencing. The *in vitro* infection studies were carried out by Dr. Nick Meyerson (Sawyer Lab), and Elena Lian (Perera Lab). Dr. Nick Meyerson and Dr. Roy Parker also provided tremendously valuable ideas and helped with the experimental design at every stage of the project. Dr. Robin Dowell has been my go-to person for all computational biology related questions. And last but not least, Dr. Sara Sawyer led the entire project since day one, she provided the necessary resources, knowledge and expertise to guide us throughout the project.

4.2 Abstract

We show that diverse bacterial, viral, and fungal infections universally upregulate the activity of 69 human genes, the mRNA transcripts of which are detectable in human saliva. First, by analyzing human genes that become upregulated in cultured human cells infected with 28 different viral, 7 different bacterial and 3 different fungal pathogens, we identified 69 human "universal response" genes which are mostly related to interferon signaling and other innate immune responses. We then recruited 32 human enrollees that were either healthy or infected with various pathogens. The abundance of these 69 mRNAs in saliva could correctly determine if the person was harboring an infection 86% of the time. Moreover, using a cohort of 48 asymptomatic but SARS-CoV-2-infected individuals, we show that the abundance of these mRNAs in saliva can correctly determine whether an individual was infected 92% of the time, even though no symptoms were present. The fact that these universal response mRNA transcripts can be monitored in saliva opens the possibility that immune system activation in response to infection could be detected non-invasively, at home, and in a pathogen-agnostic manner.

4.3 Introduction

Traditional infectious disease diagnostics focus on identifying pathogen-derived molecules such as nucleic acid or proteins. This type of diagnostic relies on a priori clinical suspicion

and the proper test(s) being conducted. At the same time, the human body is a detection device for pathogenic invaders, one that has been fine-tuned over millions of years of evolution. Certain changes in human physiology are indicative of infection in a general, pathogen-agnostic manner. Indeed, previous studies have monitored human proteins and transcripts in blood that change with infection, disease severity, and symptomatic status [74, 156–170]. However, blood sampling is predominantly carried out in a clinical setting. The ongoing COVID-19 pandemic has put a spotlight on saliva as a non-invasive biospecimen with diagnostic value, and saliva-based SARS-CoV-2 diagnostics have proven to be reliable and convenient[45, 46, 94, 128, 171].

Here, we identify 69 human "universal response" genes that are upregulated in human saliva by a broad range of bacterial, viral, and fungal pathogens. We show that the abundance of these gene transcripts in saliva can differentiate healthy from infected individuals. While some aspects of the transcriptional response to infection are specific to certain classes of pathogens, these 69 genes represent a core and universal response to infection. This is because, beyond a certain point, the addition of more datasets did not change the constitution of this group of 69: new genes were no longer added or subtracted from the set. The fact that these mRNAs can be detected in saliva opens the possibility that immune system activation in response to infection could be detected non-invasively, at home, and in a pathogen-agnostic manner.

4.4 Results

4.4.1 Identification of 69 human universal response genes

To determine human genes that are commonly upregulated in diverse infections, we first obtained a total of 71 published datasets. These datasets all profiled the transcriptional response of cultured human cells to infection. Studies involving a variety of pathogens were included (29 viruses, 7 bacteria and 3 fungi), with many of these pathogens represented by

more than one dataset (Table 4.2). Each of the 71 datasets included matched transcript sequencing for infected and mock-infected human cells, usually in multiple replicates (n = 387 replicates in all). For each dataset, raw RNA sequencing reads were retrieved from the NCBI short-read archive and analyzed as described in the Methods. We looked for genes that were upregulated in infected conditions ("+" in Figure 4.1A) compared to in mock infections ("-"). Despite the many variables in these datasets (pathogens, human cell lines, labs conducting the studies), we obtained a list of 69 genes that are consistently upregulated across the array of pathogen types tested (Figure 4.1A and genes are listed in Table 4.3). We refer to these as "universal response" genes. Therefore, while each infection triggered the expression of many human genes, these 69 genes represent a core transcriptional response that is universal. Universal response genes mainly belong to pathways related to cellular antiviral functions and type-I interferon responses (Figure 4.1B).

Several lines of evidence support the idea that these 69 genes represent a core and universal transcriptional response to infection. First, the number of universal response genes reached a non-zero asymptote of 69 genes as more studies were added to the analysis (Figure 4.1C). After reaching 69 genes, the addition of more datasets did not change the constitution of this group of 69; new genes were no longer added or subtracted from the set as datasets accumulated. Second, principal component analysis was performed on the expression data of these 69 genes in all datasets (Figure 4.1D). Of the many variables involved, the main contributor to the data variance (PC1; which explains 81.6% of the variance) cleanly separates these *in vitro* experiments by conditions of infected (triangles) or uninfected (circles). This suggested that levels of mRNAs from this group of 69 genes can differentiate infected from uninfected human cells in all cases.

We next assessed whether the abundance of these mRNAs in blinded human tissue culture samples could predict whether the cells had been infected or not. Using the 387 samples (meaning, independent experimental replicates) from the 71 *in vitro* infection datasets, we carried out cross-validation using a logistic regression model. Specifically, we first es-

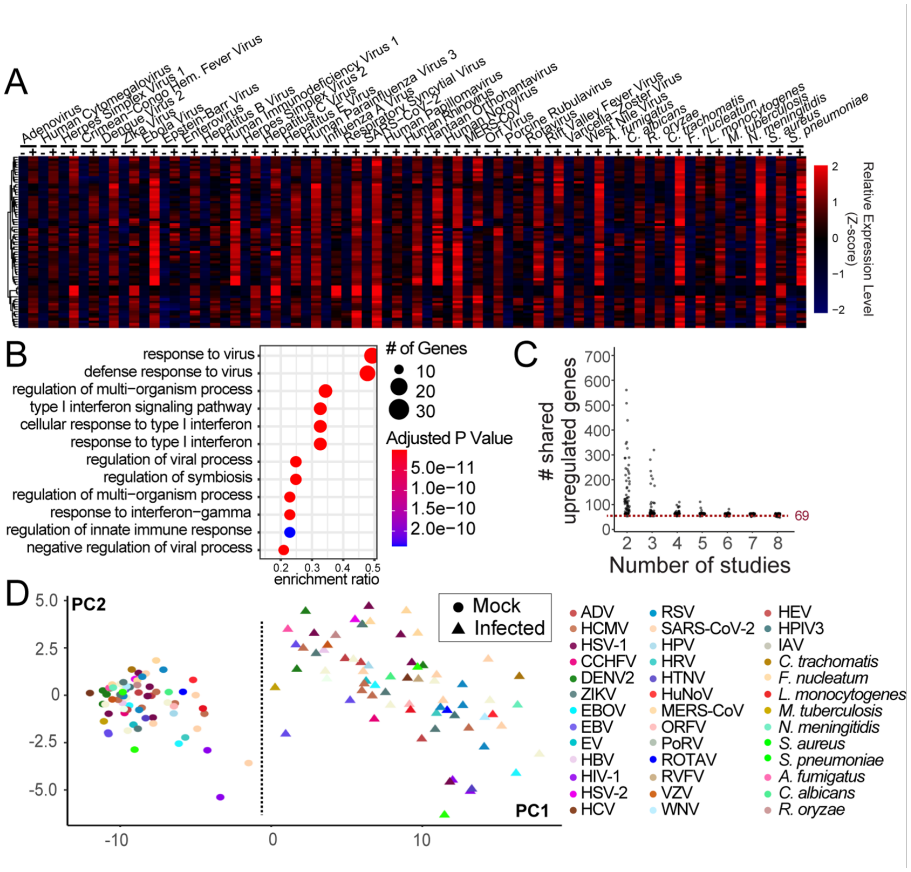


Figure 4.1: **69 human universal response genes are upregulated in a broad range of infections performed in tissue culture.** **(A)** Heatmap summarizing the observed abundance of mRNA transcripts from RNA-seq data. Each row represents one of the 69 universal response genes. Each column represents the average expression across all mock (-) or infected (+) replicates combined from all studies on a given pathogen. **(B)** A characterization of the identified universal response genes via gene ontology enrichment analysis. The X axis, enrichment ratio, is the number of observed genes divided by the number of expected genes in each GO category. The adjusted P value indicates the probability of observing the given number of genes in each category by chance. Functions related specifically to anti-viral responses are the most enriched, possibly due to an over representation of viruses within the datasets analyzed in panel A, or because innate immunity to viruses is better studied and therefore the genes involved are better annotated. **(C)** Number of commonly upregulated genes in random combinations of *in vitro* infection studies. From each of the 71 studies, we curated a list of significantly upregulated genes. We then compared these genes between randomly chosen groups of 2 – 8 studies (X axis). The X axis was truncated at 8 studies, because no new genes are added or subtracted after n=7 studies (i.e. value becomes asymptotic at n=69). **(D)** Principal component analysis of universal response gene expression data from the datasets analyzed in panel A. Mock (circles) vs. infected (triangles) samples are separated by the primary principal component (81.6% of data variance) on the x-axis. The dotted line is arbitrary but separates infected and mock samples. Virus names are defined in Table 4.2.

tablished the logistic regression classifier using the expression data of the 69 genes in 10% of the samples (much less than what is usually used in 10-fold cross-validation experiments to emphasize the predictive power), randomly selected. Next, we evaluated the predictive power of this model to classify the remaining 90% of the 387 samples as infected or not. This cross validation was repeated 10 times, and the accuracy of classification is summarized via receiver operating characteristic (ROC) curve (Figure 4.2A). Overall, the cross validation resulted in a mean area under the curve (AUC) of 0.94, also interpreted as a 94% chance of distinguishing mock from infected conditions based on the levels of these 69 mRNAs. The worst outcome of the 10 repeats had an AUC of 0.89, and the best an AUC of 0.96.

We then performed additional cross validation analyses among different types of infections (Figure 4.2B). We trained the logistic regression classifier using only fungal and bacterial samples and then classified the viral samples as infected or not. This was highly successful and yielded a ROC curve with an AUC = 0.96. We then trained the classifier using only viral and bacterial samples and then classified the fungal samples as infected or not (AUC = 1.0). Finally, we trained the classifier using a combination of viral and fungal samples, and then classified the bacterial samples as infected or not (AUC = 1.0). Collectively, this indicates that the upregulated status of these universal response genes in human cell lines can correctly identify infection status, independent of the cell and pathogen types involved. The fact that training sets on two types of pathogens can classify infections caused by a third proves that these 69 genes truly represent a universal response to infection.

4.4.2 Universal response genes are also upregulated in infected humans

We next wanted to determine if universal response genes are upregulated in infected humans. At this point, we transition from analyzing data from *in vitro* infection of human cells to the analysis of data from human biospecimens. We first took advantage of two previously published datasets from human blood, each measuring gene expression by microarray after infection. One study focused on a 34-year-old male health care worker exposed to Ebola

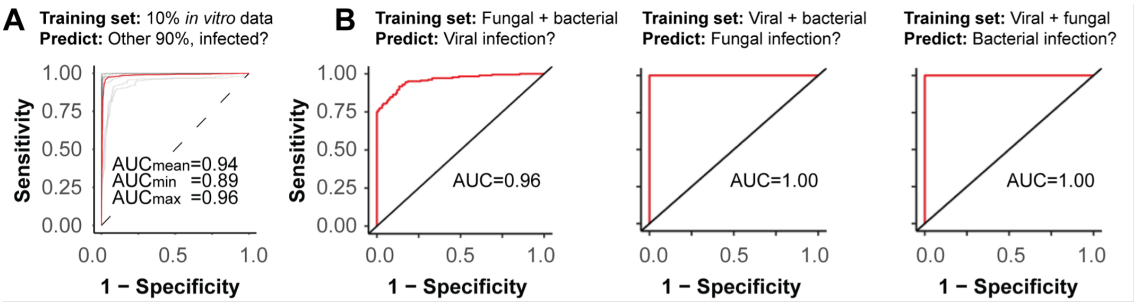


Figure 4.2: **The power of universal response genes in identifying infected human cells.** Receiver operating characteristic (ROC) curves of various linear regression models established using the expression levels of the 69 universal response genes in the 71 *in vitro* datasets used. The area under curve (AUC) is summarized in each graph. **(A)** The performance of a model trained on 10% of the samples from the 71 *in vitro* datasets. The model was them used to classify the other 90% of the samples as mock-infected or infected. The grey lines indicate each replicate of cross validation, while the red curve summarizes the average ROC curve The mean, minimum and maximum areas under curve (AUC) are indicated. **(B)** Cross validation analyses between different types of infections. In each case, the classifier was trained on infections of two types (top of graph) and used to predict whether human cells had been infected with the third type of pathogen based solely on the expression level of the 69 universal response genes.

virus in Sierra Leone during the 2013-2015 epidemic [172]. Starting 7 days after symptom onset, blood was taken from the individual daily and genome-wide mRNA expression was evaluated by microarray [173]. We extracted from this dataset the expression profiles of the universal response genes (Figure 4.3A). A vast majority of the genes are highly upregulated at day 7. Their expression trails off as the person goes through recovery, although the speed of dissipation of these signals is highly variable (a concept explored further in Figure 4.7). A few genes at the top of the panel are not upregulated at day 7, with one possibility being that their induction has already dissipated by day 7. In this individual, Ebola virus mRNA was detected between days 7-11, with the peak (Ct = 31) at day 9 [173]. From this, we can see that the strong upregulation of host universal response genes occurs at least 2 days earlier than the peak of viral load and is sustained much longer.

Another study focused on 15 individuals experimentally infected with the protist that causes malaria [174]. In this study, blood was taken every two days after experimental infection and mRNA was interrogated by microarray, until the point where individuals had detectable pathogen in the bloodstream and/or had symptoms consistent with malaria (indicated as "D" for diagnosed in Figure 4.3B). Note that protist pathogens (single-celled eukaryotes) were not represented in the 71 *in vitro* datasets from which we identified these 69 universal response genes. Nonetheless, more than half of the universal response genes (17/29) that were included on this microarray are upregulated in blood after infection. Based on these two human studies, we can say that universal response genes are also upregulated in infected humans.

4.4.3 Universal response mRNAs in saliva have diagnostic value

We next asked whether the abundance these 69 mRNAs in human saliva could classify humans as infected or not. At this point, we transitioned to analyzing human saliva samples. We first obtained saliva samples from 15 healthy (SS01-SS15; Table 4.1) and 8 infected individuals. Of the latter, three had been diagnosed with SARS-CoV-2 viral infection (enrollees

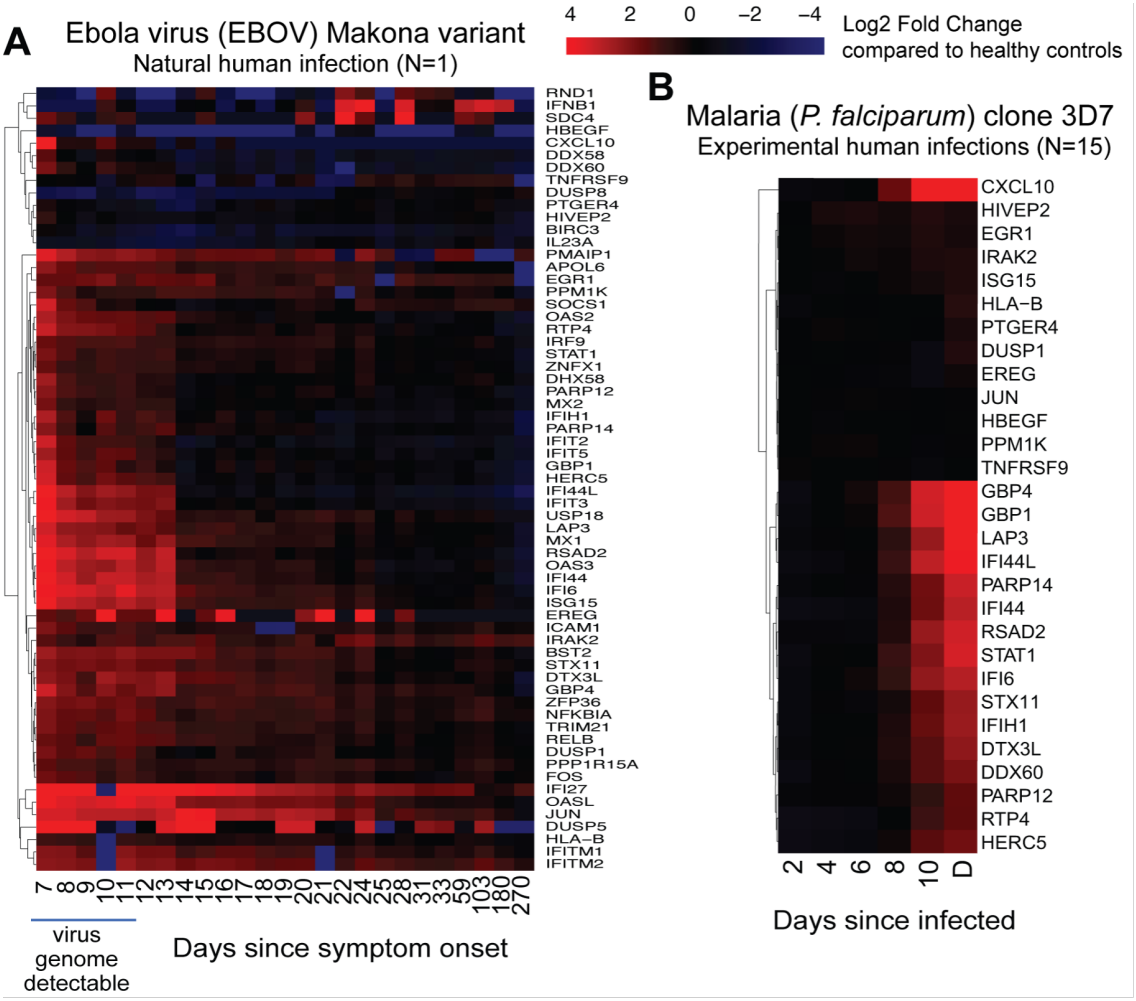


Figure 4.3: **The kinetics of transcription from universal response genes.** Heatmaps show levels of universal response mRNAs, as measured previously in transcriptome datasets from human blood samples. **(A)** This transcriptome dataset [173] was generated from a 34-year-old male health care worker exposed to Ebola virus in Sierra Leone during the 2013-2015 epidemic. Blood was taken daily starting at 7-days post-symptom onset [172]. **(B)** This transcriptome dataset is derived from 15 individuals that were experimentally infected with Plasmodium falciparum [174]. Blood was taken every two days up until the day of diagnosis ("D"). Diagnosis occurred 7.5–10.5 days post-infection, defined as the time when two of these criteria were met: positive thick blood smear, parasite density >500 parasites ml, or symptoms consistent with malaria [174]. In both studies, the transcriptome in whole blood was profiled using microarray. Only a subset of the universal response genes was included on these microarrays; hence each panel has less than 69 genes shown. The relative fold change is calculated by comparing microarray signals on the indicated day to the signal of healthy individuals from the same study (malaria N=4, Ebola N=30)

SS19-SS21), one with *Vibrio cholera* bacterial infection (SS16), one with *Staphylococcus aureus* bacterial infection (SS17), and one with varicella-zoster virus infection (VZV; SS18). Two additional saliva samples were included from apparently healthy individuals from whose saliva we were able to map reads corresponding to common respiratory pathogen genomes (SS22, CoV-NL63 seasonal coronavirus; SS23, respiratory syncytial virus (RSV)) (see Methods). Total RNA was prepared from each of these 23 human saliva samples, followed by depletion of bacterial and human ribosomal RNA. RNA with high integrity can be readily isolated from saliva (Figure 4.8). Libraries were sequenced with high-throughput short-read sequencing.

Remarkably, but consistent with our *in vitro* meta-analysis, 66 out of the 69 human universal response gene transcripts were significantly enriched in the saliva of all 8 infected individuals compared to healthy individuals (Figure 4.4A & B). In total, there were 544 genes that were significantly upregulated across all the infected individuals (light pink dots in Figure 4.4B, adjusted P-value ≤ 0.01, Fold Change ≥ 2; Table 4.4). Of these, the universal response genes are shown as dark red dots and are not necessarily the most highly upregulated transcripts.

We next tested whether the abundance of universal response mRNAs in saliva could determine if a human was harboring an infection. We carried out cross validation and found that a classifier trained on the expression levels of universal response genes in a randomly selected 10% of the *in vitro* data analyzed above (39 of the 387 experimental replicates from 71 studies), could correctly classify these 23 human saliva samples as having come from someone who is infected or healthy, just from the abundances of these mRNAs in their saliva (Figure 4.4C, Mean AUC = 0.86). Thus, this classification was made correctly 86% of the time, even with very little training data. Remarkably, the transfer learning approach (trained on *in vitro* data, then used to classify human biospecimens) only resulted in the loss of 0.08 AUC (0.94 from Figure 4.2A compared to 0.86). Therefore, the abundance of universal response mRNA in human saliva has high sensitivity and specificity to classify

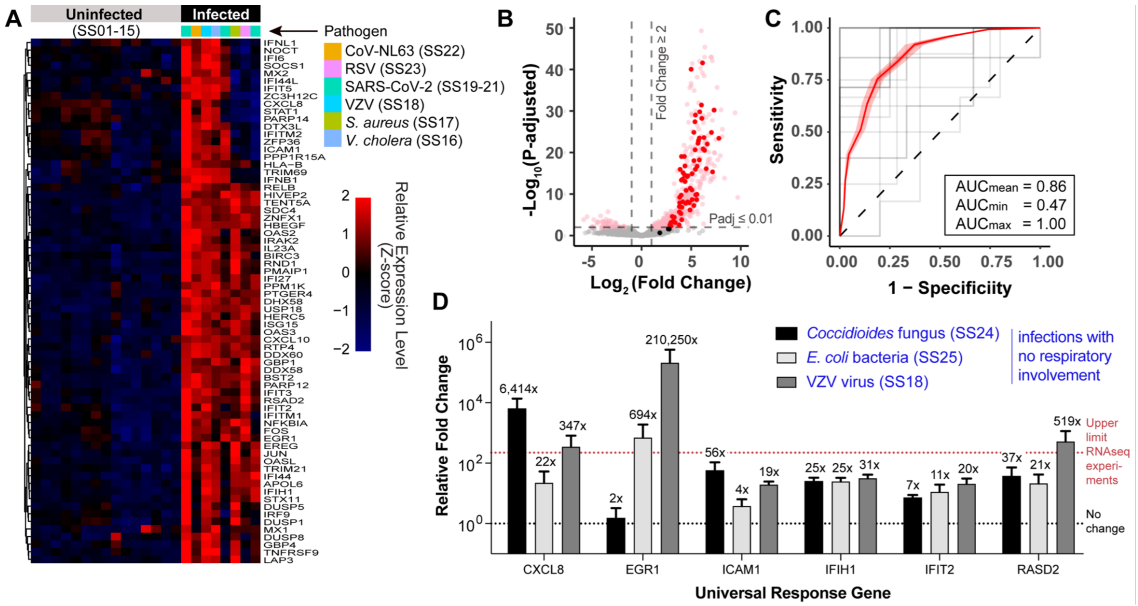


Figure 4.4: **Abundance of mRNA in human saliva can determine whether diverse infections are present in the body. (A)** Heatmap showing relative expression of each of the universal response genes (rows) in saliva, in transcripts per million (TPM) normalized to row z-score. Each column represents the saliva sample of one individual. **(B)** Volcano plot of all genes significantly upregulated in all eight infected patients compared to uninfected (DEseq2 Wald test, Fold change ≥ 2, Adjusted P-value ≤ 0.01), separated by their fold change in transcript abundance in saliva (infected vs. non-infected) and Benjamini-Hochberg adjusted p-values. The 69 universal-response genes are highlighted in dark red. **(C)** ROC curve representing the predictive power of the 69 universal response genes to distinguish healthy versus infected individuals. Logistic regression models constructed with 10% of the *in vitro* data from Figure 1, and then used to predict whether individuals SS01-SS23 were infected just based on the mRNA in the saliva. Grey lines indicate individual cross validations (N=20), the red line and shaded area indicate the average and variance from all 20 cross validations, respectively. **(D)** Total RNA from saliva from three individuals was interrogated by RT-qPCR with primers recognizing each of the universal response mRNAs shown at the bottom. To calculate the fold change of each mRNA in each infected saliva sample (shown on top of each bar), the Ct value was first normalized to the control gene, CALR, and that value was then compared pair-wise to the same value from saliva of 3 non-infected enrollees, whereafter the error bar reflects the standard error of means from the pair-wise comparison (SEM). The horizontal red line shows the highest fold-change for universal response genes in saliva observed by RNA-seq in this study, which is less sensitive.

infected humans, at least in this small set of 23 individuals.

Importantly, two of the enrollees in the previous analysis were noted to have no signs of respiratory tract involvement (SS17, SS18, Table 4.1), and some clearly had infection linked to distal sites (SS16 gastroenteritis, SS18 meningitis, Table 4.1), yet these signatures are reliably detectable in saliva. We next wanted to further confirm that universal response mRNAs can be found in saliva, even when infection is at distal sites in the body. In the next experiment, we included two additional patient saliva samples, one from an enrollee being treated for a *Coccidioides* fungal infection (SS24, Table 4.1) and another enrollee being treated for *Escherichia coli* bacterial sepsis stemming from urinary source (SS25, Table 4.1). The three enrollees in this experiment (SS18, SS24, SS25) were diagnosed with very different infections (viral, fungal, and bacterial) and were specifically noted to not have respiratory involvement in their infections (Table 4.1). We used RT-qPCR to quantify mRNA from six of the universal response genes (due to limited sample volumes) from the saliva of these enrollees. We observed from 2- to 10^5-fold upregulation of all six host mRNAs within the saliva of infected individuals compared to three healthy ones (Figure 4.4D). In summary, we can detect universal response mRNAs in human saliva, even when there is no apparent respiratory involvement. Again, a viral, bacterial, and fungal infection all lead to this noted over-abundance of universal response mRNAs in saliva.

4.4.4 Universal response transcripts in saliva can detect infection in asymptomatic SARS-CoV-2 carriers

We next asked if universal response mRNAs in saliva can identify infections even in individuals with no symptoms. During the 2020-21 academic year, the University of Colorado Boulder carried out weekly SARS-CoV-2 screening for students and staff. The screening effort enabled us to enroll university affiliates into an associated human study. All saliva samples were screened for SARS-CoV-2 by a RT-qPCR test [46, 111]. Samples were only used from enrollees who indicated the absence of any symptoms at the time of saliva donation. We

examined the levels of mRNA from universal response genes in the saliva of 48 SARS-CoV-2 positive and 20 non-infected individuals. Instead of sequencing transcripts in saliva, we developed a multiplex TaqMan RT-qPCR assay for measuring 15 of the universal response genes, along with 3 control genes (Methods, Table 4.5, Figure 4.11). These 15 genes were chosen to represent a range of expression levels and kinetics amongst the 69 total universal response genes. We observed higher levels of universal response mRNAs in the saliva of most of the SARS-CoV-2 positive individuals (Figure 4.5A). Importantly, we noticed a strong correlation between the level of mRNA observed and the saliva viral load (top of heatmap). Within saliva samples that carried high viral load, almost all had elevated level of universal response mRNAs.

The correlation between viral load and the expression of the universal response genes is highlighted by further analysis. For universal response genes, we plotted the relative fold change of mRNA in saliva against the concentration of viral genome copies in saliva (IFIT3 and IFI27 shown in Figure 4.5B, and the full set of genes analyzed is shown in Figure 4.9). For SARS-CoV-2, infectious virions are almost never recovered from individuals with viral loads below 10^6 viral copies per mL [99, 101, 103–106]. Individuals with lower viral loads are either at the beginning of infection, or on the long tail of recovery [175]. Interestingly, the mRNAs of IFIT3 and IFI27 accumulate in saliva before this point, at the transition of viral titers to above 10^4 viral copies/mL. This is consistent with a model where mRNAs from universal response genes accumulate in saliva specifically during, and possibly before, periods of acute viral replication.

Finally, to evaluate the accuracy of using universal response mRNA abundance in saliva to distinguish infected from non-infected humans, we carried out cross-validation using logistic regression models trained on RT-qPCR results from half of the enrollees in this SARS-CoV-2 study (SS33-SS100). This classifier was then used to classify the remaining human saliva samples as infected or not (Figure 4.5C). Overall, this analysis resulted in an area under curve (AUC) of 0.92 and 0.97 for detecting those infected with viral loads greater than 10^4

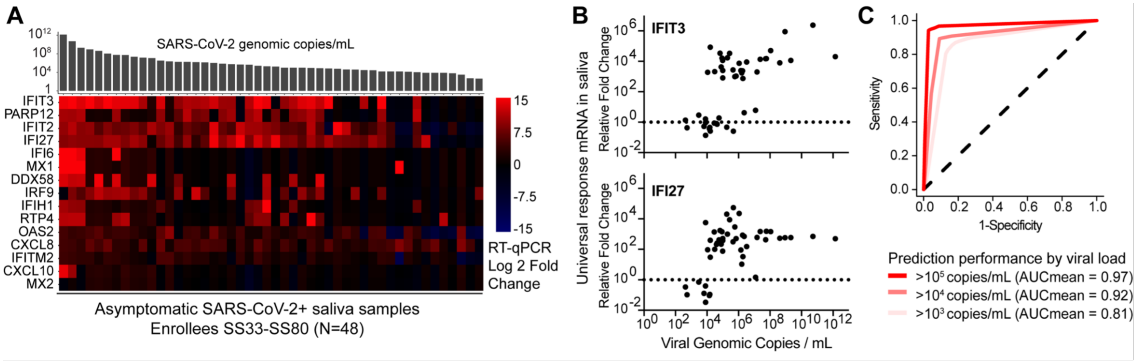


Figure 4.5: **The abundance of universal response mRNAs in human saliva can determine whether individuals are infected with SARS-CoV-2. (A)** Heatmap summarizing mRNA levels from universal response genes in the saliva of SARS-CoV-2-positive individuals. Rows represent the 15 universal response mRNAs, measured by RT-qPCR in a multiplex TaqMan assay. In columns, are individual enrollees, where the RT-qPCR value (Ct) for each mRNA in that enrollee's saliva is compared to the average Ct from 20 uninfected enrollees. The viral load in each saliva sample was measured using a separate RT-qPCR assay and is reported above the heatmap. **(B)** Scatter plots correlating the fold change of two individual human mRNAs (top) to viral loads. Each dot represents a SARS-CoV-2 infected individual. **(C)** Accuracy of universal response mRNA abundance in saliva to distinguish SARS-CoV-2-infected from uninfected individuals at different levels of viral loads. For each viral load cutoff, RT-qPCR delta Ct values from half of the SARS-CoV-2 positive samples above the indicated cutoff, along with half of the non-infected samples, were used to train the logistic regression model, while the other half was used for evaluation. The process is bootstrapped for 100 times and the average ROC curve is plotted.

genomic copies/mL and 10^5 genomic copies/mL, respectively. The evaluation again supports that the abundances of mRNAs from universal response genes, detectable in saliva, are highly reliable in predicting whether an individual is infected. This is especially true for individuals harboring viral loads consistent with the infectious phase of disease. Importantly, none of these individuals reported symptoms at the time of their saliva being collected, suggesting that the mRNAs in saliva have more predictive power over infection than even self-perceived symptoms.

Moreover, we wanted to understand the relative accumulation of universal response mRNAs in blood versus saliva. We were able to compare blood and saliva only in the case of SARS-CoV-2, so our analysis of this question will be limited. We compared total mRNA (from RNAseq) from the saliva of three hospitalized SARS-CoV-2-infected patients (SS19-SS21) to uninfected controls (SS1-SS15; Table 4.1). For each gene, this (X-axis) was plotted against the same transcript enrichment value in blood, using data from the recently published COVIDome database [176] (Figure 4.6A). Each dot is a gene, and the universal response genes are shown in red. We find that universal response transcripts (red dots), are as (or even more) detectable in saliva than in blood. This suggests universal response mRNAs can be found in both saliva and blood, at least for SARS-CoV-2, and that saliva might be the superior biological specimen in which to measure these host signatures, although more evaluation is needed.

4.4.5 Universal response mRNAs are stable in uninfected individuals over time

Finally, for universal response mRNAs in saliva to have diagnostic value, they would need to remain relatively stable in abundance over time in the absence of infection. To assess this, we enrolled 7 apparently healthy individuals (SS26-SS32; Table 4.1) who were asked to collect saliva samples daily over a period of 11 days. We then measured the level of universal response mRNAs in their saliva over the time course by RT-qPCR using the

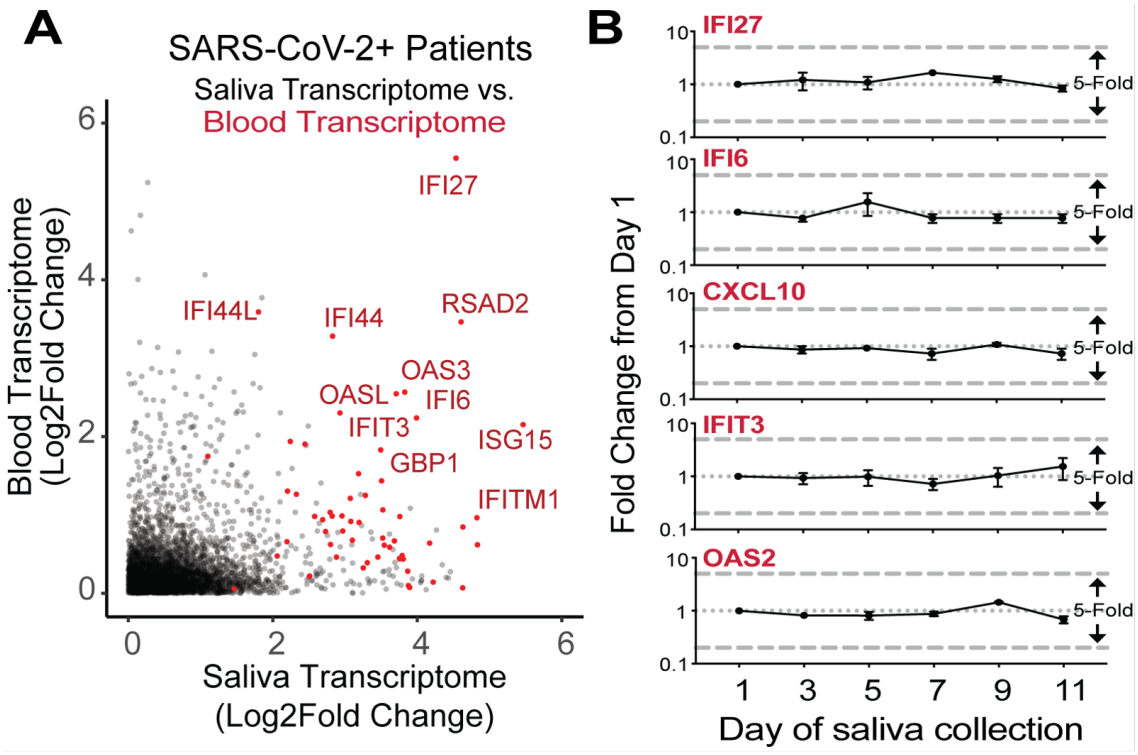


Figure 4.6: **Universal response mRNAs accumulate in blood and saliva, but not in uninfected individuals. (A)** On the X-axis, the expression levels of human mRNAs in the saliva of SARS-CoV-2+ patients (N=3, SS19-SS21, RNAseq) were compared that of uninfected control individuals (N=15, SS1-SS15). The plot shows only genes with fold change ≥ 1. On the Y axis is the similar analysis, performed in the blood in individuals from a different SARS-CoV-2 cohort[176] . The universal response genes are highlighted by red dots, with the most highly enriched labeled with gene symbols. **(B)** To determine the extent of mRNA variation from day to day in human saliva samples, 7 apparently healthy individuals (SS26-SS32) were asked to collect saliva daily for 11 days. Total RNA was isolated from each sample and used as a template for a multiplex TaqMan assay measuring the levels of 15 universal response genes. Five of the universal response genes are shown, and the remainder are shown in Figure S4. For each of the 7 enrollees, their Ct value for each gene was converted to fold change by normalizing it to the Ct value of RPP30, and then again to the abundance of mRNA measured at Day 1. Error bars represent the SEM of 7 individuals.

multiplex TaqMan assay described above. The expression levels of the universal response genes remained remarkably stable over time (five genes shown in Figure 4.6B, the full set in Figure 4.10). When compared to day 1, transcript abundance in saliva changed no more than 5-fold in subsequent days. Thus, universal response mRNAs are remarkably steady in the saliva of healthy individuals.

4.5 Discussion

Here, we identify 69 human "universal response" genes that are upregulated by a broad range of bacterial, viral, and fungal pathogens. Even when infection resides in distal sites in the body, the mRNAs produced in this universal response are measurable in human saliva. By assessing the abundance of these mRNAs in saliva, we were able to correctly determine whether a person harbors infection more than 85% of the time. This is true even in the absence of perceived symptoms, meaning that in some cases the enrollees themselves didn't even know they were sick. Therefore, the monitoring of these mRNAs in saliva could be a platform for detecting infection in the body, even as a screening tool for healthy individuals.

It is striking that there is a core transcriptional response that is triggered by all tested pathogens. Many studies have explored the host gene response to infection, including the 71 studies that we used in the first step of our own study here (listed in Table 4.2), or to specific cytokines like interferon[177, 178]. Yet there have been far fewer studies that have looked at commonalities in gene induction by cells infected with different pathogens, and typically these have compared just a few pathogen types [179, 180]. By integrating results from many datasets from a broad range of pathogen types, we identified an asymptotic number of universal response genes (n=69). Importantly, no new genes were added or subtracted from this list once we surpassed a certain number of datasets analyzed (after n=7). Thus, we identified the connecting signature[181] that underlies infection, across a broad range of pathogens.

Importantly, universal response mRNAs are detectable in saliva of infected individuals,

regardless of the location of infection. There are two hypotheses to explain why these mRNAs are found in saliva. First, free mRNA, or mRNA encapsulated in dead cells or exosomes, might be entering the mouth. This might be occurring for the purpose of targeting these structures for elimination from the body via the gastrointestinal tract. In a second model, interferon and other cytokines produced by a distal infection may be entering the oral cavity and stimulating cells there to execute the transcriptional response that we are measuring [182]. In other words, the mRNA we observe in saliva could be produced or even propagated locally in the mouth. Ultimately, we cannot differentiate between the possibilities that these mRNAs are transported to the mouth or that they are produced in the mouth. Regardless, this study highlights the diagnostic value of saliva beyond its current limited use in diagnosing SARS-CoV-2, oral cancers, and Sjorgen syndrome[183]. Future work should focus on parallel characterization of transcriptomics, proteomics, and metabolomics in both blood and saliva so that we can grasp the full scale of systemic response to infection in different bodily fluids.

This is but the first iteration of this diagnostic approach in saliva, and there are obvious improvements that could be made. First, it is unclear whether all 69 genes together bear the most power for detecting infection in the body, or whether a smaller subset of them would carry as much power to do so. The Golub lab has shown that there is an underlying connectivity map between transcripts of every gene in the human genome [145, 181]. They have identified a core set of 1000 human genes that, when measured, are sufficient for inferring (81% accuracy) the entire transcriptional response of cells to a certain drug that is administered [181]. Likewise, the expression levels of a few of the universal response genes might be informative of the expression levels of others of them. More information is also needed on the kinetics of production and destruction of the universal response mRNAs during the time course of an infection (but see Figure 4.7). This information might also refine the set of mRNAs designed to optimally diagnose someone as infected. Finally, for universal response mRNAs to be used for diagnosis of infection, it will be crucial to determine the role of human genetic diversity in this transcriptional response. Previous reports have shown

that people with different ancestry elicit differential immune responses [184]. On some level, there may be a core transcriptional response to infection that is shared by all humans. This is because, as shown by the Palmarini group, there is a core set of genes upregulated by interferon that is conserved even across a broad range of mammalian species [185]. A final interesting follow-up question is whether it would be therapeutically useful to trigger this universal transcriptional response with an agonist in the case of untreatable infections and, if so, how that might be done.

4.6 Material and Methods

4.6.1 Meta-analysis of NCBI SRA transcriptomics datasets

We carried out meta-analysis of RNA-seq datasets publicly available at the NCBI SRA database. Our criteria for choosing datasets where that human cells in culture were infected with a bacterial, viral, or fungal pathogen, and then the cellular transcriptome was sequenced along with that in a mock-infected control. We obtained a total of 71 relevant *in vitro* infection datasets. From these datasets, raw RNA sequencing reads in FASTQ format were downloaded, trimmed using BBDuk (BBMap v38.05)[186] and mapped using HISAT2 v2.1.0[187] to human genome assembly hg38. Using NCBI RefSeq genome annotation, we then counted the mapped reads assigned to gene or transcripts using FeatureCount (Subread v1.6.2)[188].

First, we looked for genes that were upregulated in each infected dataset versus its matched mock infection. For each individual dataset, the infected replicates were compared to the corresponding mock replicates via the DESeq2 Wald test (v3.1.3)[10], from which the fold change and Benjamini-Hochberg adjusted p-values were obtained. Correction for multiple testing was performed throughout. Next, we looked for the subset of these genes that was statistically enriched in infected datasets overall. DESeq2 results from individual datasets were ranked and combined based on the magnitude and consistency of upregulation

across the datasets. Specifically, the gene rank, r_g is assigned to each individual dataset following the formula:

$$r_g = Rank(-log_{10}(Pval_{adj}) \times fold\_change)$$

Next, to determine which gene is consistently upregulated across different studies, the rank is combined via rank sum statistics. With n studies, the rank sum for each gene, g, is calculated as:

$$RS_g = \sum_i r_{g,i}$$

Hence, each gene is sorted based on the RS_g. We then filtered the gene list based on the within-study adjusted p-value and required that the gene to be significant (padj$\leq$0.05) in 90% of the datasets. As the result, we obtained 69 universal response genes ranked by the statistical significance comparing infected vs. mock groups and by the consistency across datasets.

4.6.2 Cross validation using logistic regression models

To evaluate the predictive power of the universal response genes in differentiating infected/uninfected conditions in *in vitro* and *in vivo* RNA-seq datasets, we extracted library size-normalized read counts in transcript per million format for each sequencing replicate. We next separated the datasets into training and prediction set. Specifically, 10% of randomly selected sequencing replicates from *in vitro* infection datasets were used to construct the logistic regression model using R package stats (v 3.6.2) glm function with specification of "family=binomial(link='logit')" for binomial logistic regression. The remaining 90% of sequencing replicates were used as the predict set for evaluation. In the case of *in vivo* saliva sequencing replicates, the entire dataset was used for prediction. R package ROCR (v1.0.11)57 was used to generate the ROC curves based on the prediction outcome.

For evaluating the predictive power of universal response genes as measured by the TaqMan RT-qPCR assay on SARS-CoV-2 infected/uninfected saliva samples, the relative

fold change was calculated by first normalizing the raw Ct values to the corresponding control gene Ct (RPP30) and then comparing to the average normalized Ct of all uninfected individuals. The relative fold change values for each individual were then used for cross validation via logistic regression model. Specifically, half of infected individuals above the said viral load threshold along with half of the uninfected individuals are used as the training set, while the remaining half was used for prediction. The methods for constructing the logistic regression model and for evaluating performance via ROC are same as above.

4.6.3 Human saliva sample collection, handling, and RNA preparation

Samples SS4, SS5, SS12-SS21, SS24 and SS25 were collected under protocol 17-0562 (U. Colorado Anschutz Medical School; PI Poeschla), where adult participants were consented verbally and donated up to 5mL of whole saliva and/or 50mL whole blood per visit with no more than two visits per week and no more than 500mL blood volume drawn per patient. Saliva was collected into Oragene saliva collection kit (DNA Genotek CP-100). The saliva is mixed with the stabilization solution in the collection kit and stored at room temperature for no longer than 2 weeks before being processed for RNA purification. Blood collected from patients with confirmed or suspected infection did not exceed the lesser of 50mL or 3mL per kilogram in an eight-week period. Diagnosis of these individuals was provided in the form of clinical notes.

Saliva samples from individuals SS1-SS3, SS6-SS11, SS22, and SS23 were collected under protocol 19-0696 (U. Colorado Boulder, PI Sawyer), where anonymous adults verbally consented and donated up to 2mL of whole saliva. Saliva was collected into Oragene saliva collection kit as mentioned above. For these individuals, infection status was later determined by in silico metagenomic detection using GOTTCHA (v1.0b)[189] using the RNAseq reads (additional RNAseq sample preparation and analysis described below). We were able to detect sequencing reads mapping to CoV-NL63 or RSV genomes from the saliva of individual SS22 and SS23, respectively, so they were presumably infected with these pathogens at the

time of saliva donation.

Saliva samples for apparently healthy individuals over a daily time course (SS26-SS32) were collected under a COVID-19-related sub-study of protocol 19-0696 (U. Colorado Boulder, PI Sawyer), where adult participants consented verbally and donated up to 2mL of whole saliva per day of participation up to a total of 28mL of whole saliva. The saliva was collected into Oragene saliva collection kit as mentioned above.

To purify RNA from saliva samples collected in Oragene saliva collection kit, we used 1 mL saliva 1:1 diluted in stabilization solution and followed the manufacturer recommended protocol by DNA Genotek to precipitate the nucleic acid. The RNA is further DNase-digested using Turbo DNase (Invitrogen #AM2238) and cleaned up using RNA clean-up and concentration micro-elute kit (Norgen #61000). The purified RNA is used for RT-qPCR or processed further for RNA-seq.

To prepare the total RNA for sequencing, we first spiked in ERCC RNA spike-in mix (ThermoFisher #4456740) into the saliva total RNA for downstream normalization. We depleted bacterial ribosomal RNA using pan-bacterial riboPOOL kit (siTOOLS #026). We then prepared the RNA for total RNA sequencing using KAPA RNA HyperPrep kit with RiboErase to remove human rRNA (Roche #KK8560). Finally, the saliva total RNA libraries were sequenced in 150 bp pair-end format using NovaSeq 6000 (Illumina) at the depth of 30 million reads.

Saliva samples for SARS-CoV-2-infected individuals (SS33-SS80), and matched SARS-CoV-2-negative individuals (SS81-SS100) were collected under protocol 20-0417 (U. Colorado Boulder, PI Sawyer), where adult participants 17 years of age or older (under a Waiver of Parental Consent) provided written consent. These samples were collected and tested for the SARS-CoV-2 virus during our campus COVID-19 testing initiative 24,27 during the Fall 2020, Spring 2021, and Summer 2021 semesters. As part of this campus testing operation, university affiliates were asked to fill out a questionnaire to confirm that they did not present any symptoms consistent with COVID-19 at the time of sample donation, and to collect no

less than 0.5 mL of saliva into a 5-mL screw-top collection tube. Saliva samples were heated at 95 °C for 30 min on site to inactivate the viral particles for safer handling, and then placed on ice or at 4 °C before being transported to the testing laboratory for RT-qPCR-based SARS-CoV-2 testing performed on the same day. Samples were then kept in -80 °C until RNA preparation. The total RNA of the remaining saliva samples was then purified using TRIzol LS reagent (ThermoFisher #10296028) followed by GeneJET RNA cleanup and concentration kit (ThermoFisher #K0841). The purified total RNA was used for RT-qPCR following the steps described below.

Additional saliva samples for general assay development were collected under protocol 20-0068 (U. Colorado Boulder, PI Sawyer), where anonymous adult participants were verbally consented and donated up to 2mL of whole saliva for use as a reagent in optimization and limit of detection experiments.

4.6.4 Analysis of high-throughput transcriptomics data from human saliva samples

To profile human transcriptomic changes in human saliva samples, raw RNA sequencing reads in FASTQ format were obtained, trimmed using BBDuk (BBTools v38.05), and mapped using HISAT2 v2.1.0 to human genome assembly hg38 along with ERCC spike-in sequence reference. Using NCBI RefSeq genome annotation (GRCh38.p13), we then counted the mapped reads assigned to gene or transcripts using FeatureCount (Subread v1.6.2). Read counts was first normalized using R package RUVseq (v1.28.0)[190] to account for library size factors based on the ERCC spike-in counts. Individual samples were then separated into infected and non-infected groups and the differential expression of genes were determined via DESeq2 (v3.1.3) Wald test[10], from which the fold change and Benjamini-Hochberg adjusted p-values were obtained.

4.6.5 RT-qPCR analysis of universal response genes in human saliva

For initial RT-qPCR validation on 3 clinically diagnosed and 3 uninfected samples (Figure 4.4D), 2 µL of saliva total RNA was first reverse transcribed to cDNA using poly-dT primers with the SuperScript IV first-strand synthesis system (Invitrogen #18091050). The saliva cDNA was diluted 1:20, and 5 uL of the cDNA dilution was used for each qPCR reaction including 10 µL PowerUp SYBR Green master mix (AppliedBiosystems # A25741), 500 nM forward and reverse primers (table below), and nuclease free water. The qPCR assay was carried out on QuantStudio3 real-time PCR system (ThermoFisher) consisting of a UDG activation step (50°C for 2 min, 95°C for 2 min), 40 cycles of PCR stage (95 °C for 15 s, 60 °C for 60 s, with a 1.6 °C/s ramp-up and ramp-down rate), followed by a melt curve stage (95 °C for 15 s, 60 °C for 60 s, slow ramp-up to 95 °C at 0.15 °C/s). The cycle threshold (Ct) values were used to calculate relative fold change using delta delta Ct method.

Gene Name	Forward Primer Sequence (5'-3')	Reverse Primer Sequence (5'-3')
CALR	TCCCGATCCCAGTATCTATGCC	TCTCTGCTGCCTTTGTTACGC
CXCL8	CCAGGAAGAAACCACCGGAA	CTTGGCAAAACTGCACCTTCAC
EGR1	ACTACCCTAAGCTGGAGGAGA	AGGAAAAGACTCTGCGGTCA
ICAM1	GCAACCTCAGCCTCGCTAT	GGAGTCCAGTACACGGTGAG
IFIH1	ACAGCTTCACCTGGTGTTGGA	ATGGCAAACTTCTTGCATGGCT
IFIT2	CCCTGCCGAACAGCTGAGAA	AGTTGCCGTAGGCTGCTCTC
RSAD2	GTTGGTGAGGTTCTGCAAAGTAGAGTTGCG	TAAGGTAGGAGTCTTTCATCTTCTGGTTAG

Multiplex RT-qPCR analysis for the quantitative detection of human gene transcripts was carried out using customized and multiplexed TaqMan primer and probe mixes (Supplemental Table 4.4). Understanding that the contamination of genomic DNA often introduces quantification bias when measuring host gene expression, we explicitly designed primers that span exon junctions and limit the assay elongation time so that only the host RNA is reverse transcribed and amplified. As each transcript varies in its expression magnitude, we assigned genes into multiplex groups based on similar expression magnitudes observed in the meta-analysis of *in vivo* datasets and in human saliva. This minimizes competition of amplification reagents. Specifically, to determine the host gene expression levels, 1.5 µL of customized TaqMan multiplex probes were mixed with 5 µL 4X TaqPath 1-step multiplex

master mix (ThermoFisher # A28526), 5 µL of saliva total RNA, and 8.5 µL of nuclease free water. The RT-qPCR assay was carried out on QuantStudio3 Real-time PCR system (ThermoFisher) consisting of a reverse transcription stage (25°C for 2 min, 50°C for 15 min, 95°C for 2 min) followed by 45 cycles of PCR stage (95 °C for 3 s, 55 °C for 30 s, with a 1.6 °C/s ramp-up and ramp-down rate). The cycle threshold (Ct) values were used to calculate relative fold change using delta delta Ct method. For the choice of internal control genes, we combined the meta-analysis (Figure 4.1; cell culture experiments) and the saliva RNA-seq datasets (Figure 4.2; human samples) to select genes for which the expression level remained most constant and abundant across the various conditions inherent to these experiments.

4.6.6 Infection of A549 cells with influenza A virus

For influenza A virus infection, human lung epithelial cells (A549s) where plated at a concentration of 1x10^6 cells/well in a 6-well plate. The next day, the cells were infected with influenza A virus (Influenza A/Udorn/307/72) at an MOI=0.1 in serum-free media containing 1.0% bovine serum albumin. After 1 hour incubation, the inoculum was removed and replaced with growth media containing 1 ug/mL of N-acetylated trypsin. 24 hours post-infection, total RNA was harvest using QIAGEN RNeasy Mini kit (QIAGEN #74104).

4.6.7 Infection of Huh7 cells with SARS-CoV-2

Human Hepatoma (Huh7) cells (gift from Charles Rice, Rockefeller University) were grown in 1X DMEM (ThermoFisher cat. no. 12500062) supplemented with 2 mM L-glutamine (Hyclone cat. no. H30034.01), non-essential amino acids (Hyclone cat. no. SH30238.01), and 10% heat inactivated Fetal Bovine Serum (FBS) (Atlas Biologicals cat. no. EF-0500-A). The virus strain used for the assay was SARS-CoV2, USA WA 01/2020, passage 3. Virus stocks were obtained from BEI Resources and amplified in Vero E6 cells to Passage 3 (P3) with a titer of 5.5×10^5 PFU/mL. Cells were resuspended to 6.0×10^5 cells/mL in 10% DMEM and seeded at 2 mL/well in 6-well plates. The plates were then

incubated for approximately 24 hours (h) at 37°C, 5% CO2 for cells to adhere prior to infection. Cell were infected with SARS-CoV-2 at an MOI of 0.01. Samples were harvested at 0, 2, 4, 8, 12, 24, and 48 hours post infection in 200 μl TRIzol reagent for RNA extractions following the manufacture's protocol.

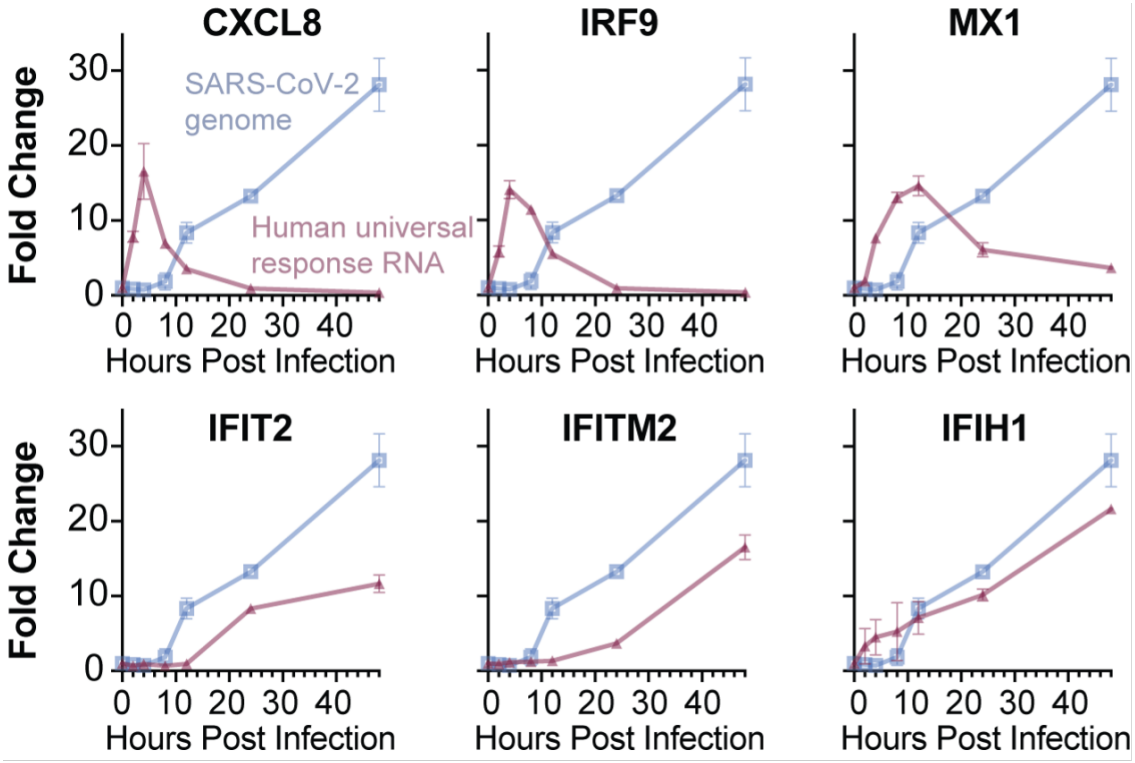


Figure 4.7: **Universal response genes are up- and down-regulated with different kinetics upon infection.** Huh7 human liver cells were infected with SARS-CoV-2 at MOI of 0.01 over a time course of 48 hours. Total RNA was harvested 0, 2, 4, 8, 12, 24, and 48 hours post infection. The fold changes of six universal response mRNAs (top of each graph; red data line) and of the SARS-CoV-2 genome (blue data line) were measured by RT-qPCR. Error bars represent the SEM of 3 biological replicates. Ct value is converted to fold change by normalizing the Ct value to the Ct value of RPP30, and then normalized again to the abundance of mRNA measured in a mock infection. Some universal response genes (CXCL8, IRF9, MX1) are upregulated in the early time points of the infection and then rapidly downregulated within the first 24 hours. This is quite interesting, since this is a low-MOI spreading infection and new cells are constantly getting infected. This would be consistent with a pulse of activity that is then quickly downregulated by a feedback loop. On the other hand, the upregulation of other universal response genes (such as the classical type-I interferon inducible genes, IFIT2, IFITM2, and IFIH1), starts later and increases steadily along with viral genome replication. This result suggests that the abundance of mRNA from any specific universal response gene will depend on the timepoint during infection, even in situations of spreading infections as would be the case in the human body.

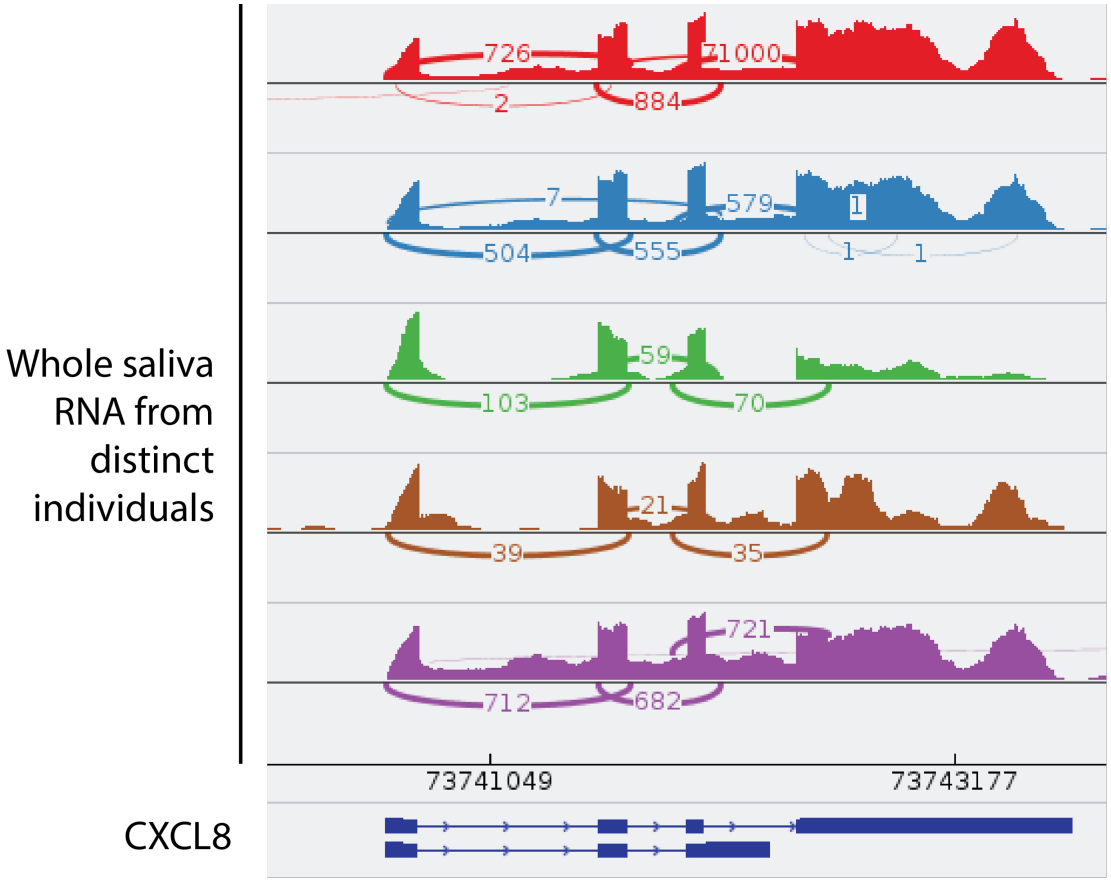


Figure 4.8: **mRNA structure is preserved in human saliva samples** Sashimi plot indicating mRNA structure is preserved during the saliva sample processing and collection, so that the exon regions are preferentially sequenced over the introns. Shown here are saliva samples from 5 individuals, CXCL8 gene is selected as the example.

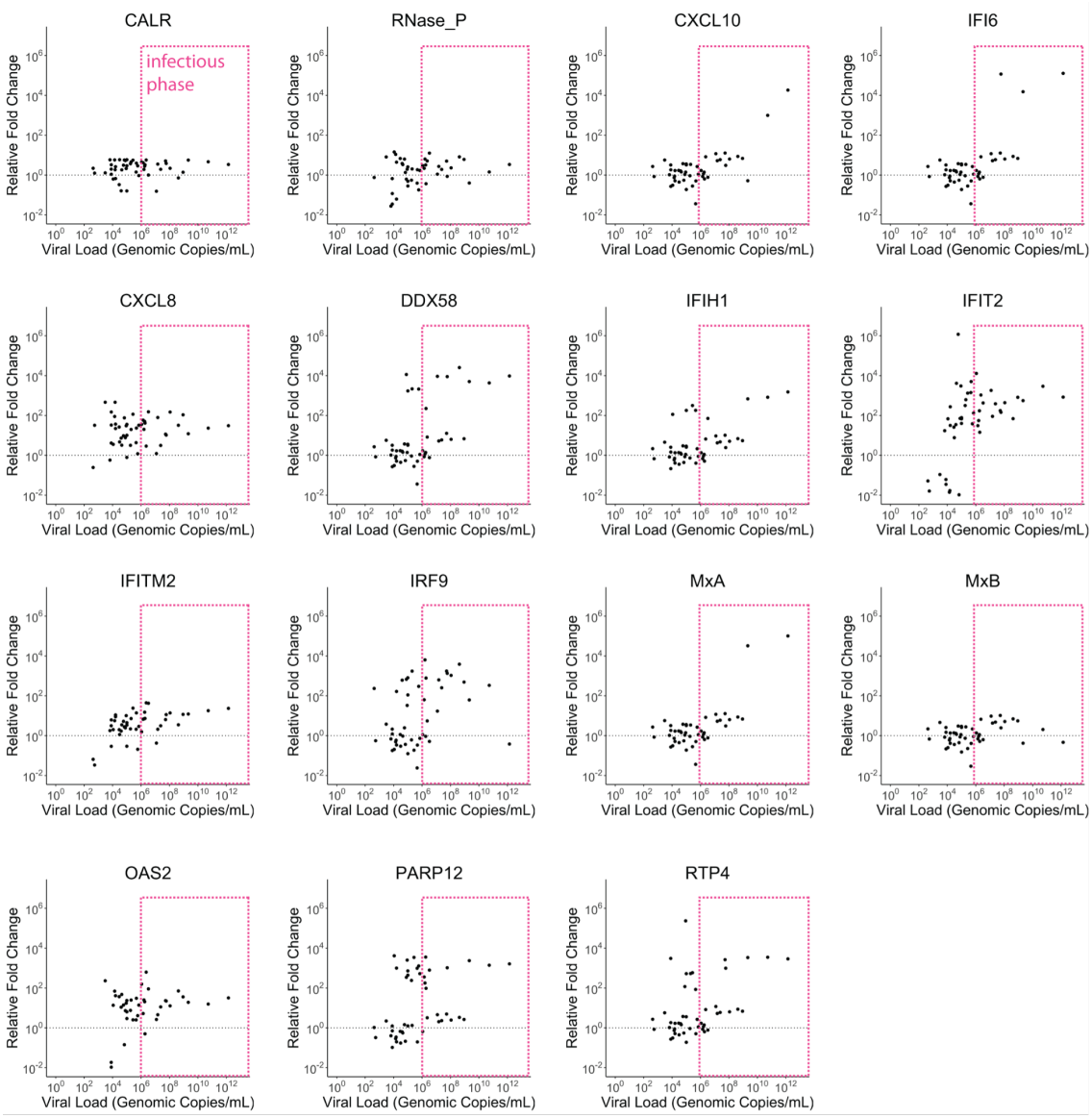


Figure 4.9: **Abundance of universal response mRNA in human saliva correlates with relative viral load in saliva samples of SARS-CoV-2+ individuals** Scatter plot where each dot represents a saliva sample from a SARS-CoV-2-positive individual. The X-axis corresponds to SARS-CoV-2 viral load, determined by RT-qPCR. The Y-axis shows the relative fold change of the human mRNA noted at the top of the graph, determined by RT-qPCR. Each measurement of human mRNA was compared to the average of the same measurement from the saliva of 20 uninfected samples, to calculate the relative fold change that is shown. The horizontal dashed line indicates the fold change of 1. A pink box shows the range of viral loads above which infectious virions can be isolated from humans (usually considered above 10^6 viral copies/mL [99, 101–105, 191]). This is consistent with a model where mRNAs from universal response genes accumulate in saliva specifically during periods of acute viral replication and the infectious phase of disease.

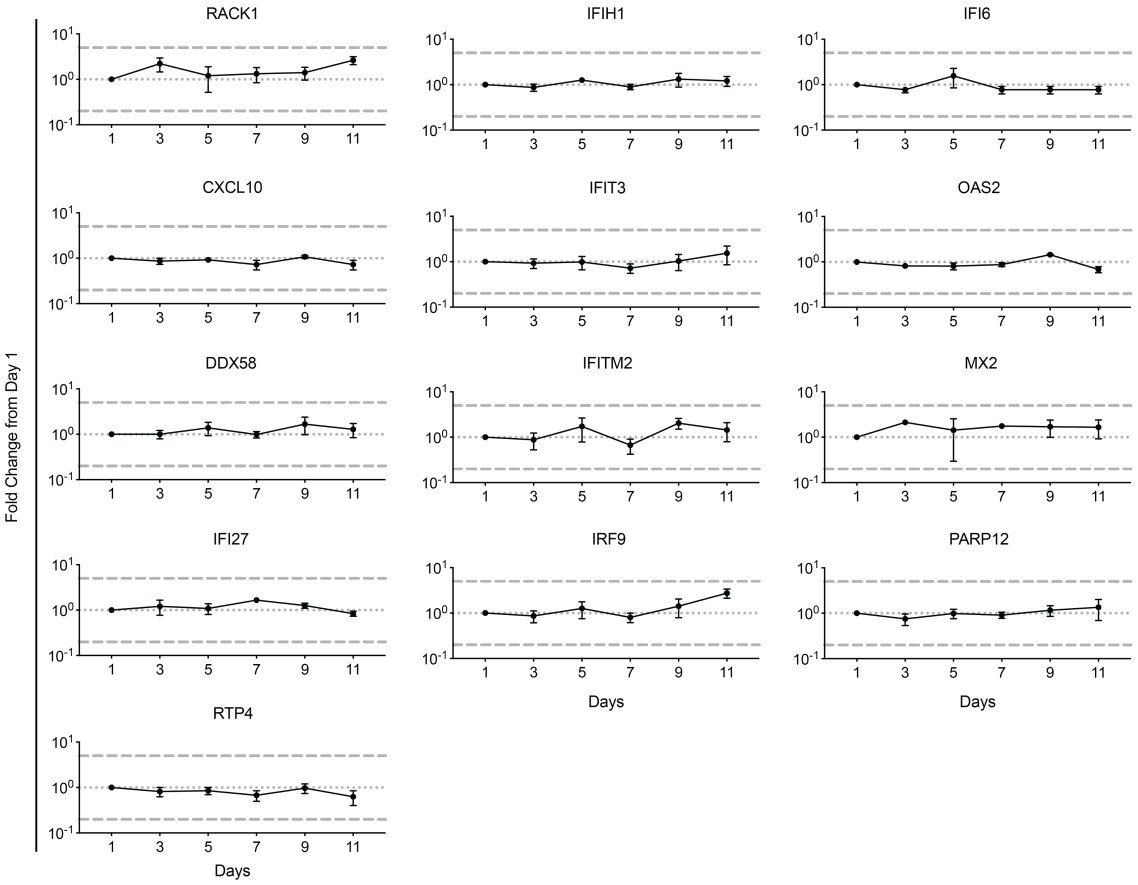


Figure 4.10: **Relative fold change of the control genes and the universal response genes over time in healthy human saliva** To determine the extent of mRNA variation from day to day in human saliva samples, 7 individuals (SS26-SS32) were asked to collect saliva on daily basis over a period of 11 days. Total RNA was isolated from each sample and used as a template in the multiplex TaqMan assay described. Shown here are the 1 control gene (RACK1) and 12 universal response genes (IFIH1, IFI6, CXCL10, IFIT3, OAS2, DDX58, IFITM2, MX2, IFI27, IRF9, PARP12 and RTP4) quantified. Error bars represent the SEM of 7 individuals. In all panels, Ct value is converted to fold change by normalizing the Ct value to the Ct value of RPP30, and then normalized again to the abundance of mRNA measured on Day 1 for each individual.

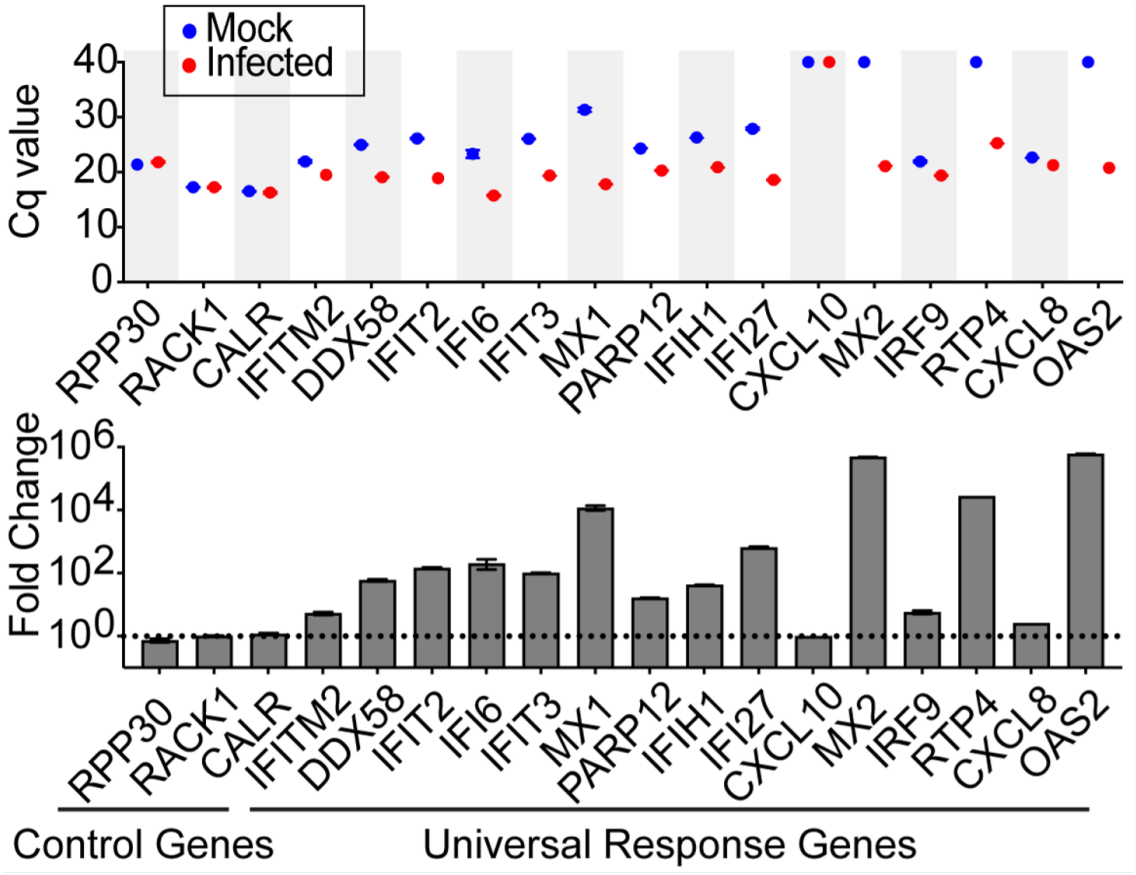


Figure 4.11: **Optimization of TaqMan assay in cells infected with influenza A virus** A549 human lung cells were infected with Influenza A virus at multiplicity of infection (MOI) of 0.1 for 24 hours. Total RNA was harvested from the cells and 100 ng was used as template in the multiplex TaqMan assay described. To demonstrate the dynamic range and the signal consistency, the raw Ct values are shown in the top panel, and the resulting fold changes are shown in the bottom panel. The error bar indicates the SEM from 2 biological replicates. Ct value is converted to fold change by normalizing the Ct value to the Ct value of RPP30, and then normalized again to the abundance of mRNA measured in a mock infection.

Table 4.1: **Human saliva samples used in this study**

Enrollee	Collection date	Diagnosis / infectious agent with clinical notes where available	Respiratory assessment from clinical notes where available	Study Site[1]
SS01-15	March-Dec. 2019	15 apparently healthy enrollees		1,2
SS16	September 2019	Patient with gastroenteritis caused by ***Vibrio cholera* bacteria.** Received one dose of Cipro and ceftriaxone before saliva sample taken.		1
SS17	September 2019	Patient with Methicillin-resistant ***Staphylococcus aureus* (MRSA)** bacteremia, C5-C7 osteomyelitis/discitis, prevertebral fluid concerning for abscess.	No erythema noted in mouth/pharynx. Pulmonary effort is normal and breath sounds normal.	
SS18	September 2019	Patient with **varicella-zoster virus (VZV)** meningitis. Herpes Zoster. Meningitis, involving left V1-V2 dermatome without ocular involvement.	Lungs CTAB, no wheezes, rales or rhonchi, normal effort.	
SS19 - 21	May 2020	3 patients being treated for **SARS-CoV-2** infection. Saliva samples were taken 4-7 days after diagnosis.		
SS22	January 2020	University affiliate whose saliva contained RNAseq reads mapping to **seasonal coronavirus CoV-NL63**.		2
SS23	February 2020	University affiliate whose saliva contained RNAseq reads mapping to **respiratory syncytial virus (RSV)**.		
SS24	February 2019	Patient with disseminated Coccidioidomycosis (Valley Fever; infection with the **fungus *Coccidioides***). Face and extremity annular lesions over last months.	No shortness of breath or cough. No oral ulcerations or nasal discharge. No evidence of pulmonary disease on imaging.	1
SS25	December 2019	Patient undergoing sepsis, likely 2.2 odds ratio pyelonephritis by ***Escherichia coli***. Most likely etiology is urinary source.	Lungs clear sounding, no rales, rhonchi or wheezes.	
SS26 - 32	May – Aug. 2020	7 apparently healthy individuals who provided saliva samples daily for 11 days		2
SS33 - 80	August – Dec. 2020	48 **SARS-CoV-20** diagnosed individuals reporting no symptoms at time of saliva collection (university affiliates)		
SS81 - 100		20 covid-negative and apparently healthy university affiliates		

[1] *Study sites: 1) University of Colorado Anschutz Medical School, 2) University of Colorado Boulder*

Table 4.2: **Transcriptomics datasets used for the discovery of human universal response genes**

SRP Index	Human cell line	Pathogen	Virus, Bacteria, Fungus	Hours Post-Infection	Sequencing Data Type
SRP044763	IMR90	Adenovirus	Virus	24	mRNA
SRP163661	MRC5	Adenovirus	Virus	24	Total
SRP202003	HepG2	Crimean-Congo hemorrhagic fever virus	Virus	72	Total
SRP078309	A549	Dengue Virus 2	Virus	36	Total
SRP130978	HUH751	Dengue Virus 2	Virus	NA	Total
SRP132737	Huh7	Dengue Virus 2	Virus	18	Total
SRP188490	HEK293	Dengue Virus 2	Virus	18	Total
SRP060253	AGS	Ebola Virus	Virus	NA	Total
SRP101856	DC	Ebola Virus	Virus	24	Total
SRP111145	ARPE19	Ebola Virus	Virus	24	Total
SRP255890	B Cell	Ebola Virus	Virus	NA	Total
SRP272684	B Cell Lymphoma	Ebola Virus	Virus	24	Total
SRP131318	Rhabdomyosarcoma	Enterovirus	Virus	6	Total
SRP212863	HUVEC	Hantaan Orthohantavirus	Virus	72	Total
SRP158789	HepG2	Hepatitis B Virus	Virus	72	Total
SRP187206	HUH751	Hepatitis C Virus	Virus	148	Total
SRP091538	HepG2	Hepatitis E Virus	Virus	120	Total
SRP117344	KMB17	Herpes Simplex Virus 1	Virus	48	Total
SRP154536	HEK293	Herpes Simplex Virus 1	Virus	4	Total
SRP163661	MRC5	Herpes Simplex Virus 1	Virus	9	Total
SRP177947	THP1	Herpes Simplex Virus 1	Virus	24	Total
SRP189489	HFF	Herpes Simplex Virus 1	Virus	8	Total
SRP065236	HFF	Herpes Simplex Virus 2	Virus	8	Total
SRP065236	EC	Human Cytomegalovirus	Virus	48	Total
SRP065236	HFF	Human Cytomegalovirus	Virus	48	Total
SRP065236	NPC	Human Cytomegalovirus	Virus	48	Total
SRP163661	MRC5	Human Cytomegalovirus	Virus	48	Total
SRP266618	NTT	Human Cytomegalovirus	Virus	24	Total
SRP065236	CD4+ T Cell	Human Immunodeficiency Virus 1	Virus	120	Total
SRP155217	CD4+ T Cell	Human Immunodeficiency Virus 1	Virus	72	Total
SRP155822	Ileum organoid	Human Norovirus	Virus	48	Total
SRP223234	HFK	Human Papillomavirus	Virus	NA	Total
SRP253951	A549	Human Parainfluenza Virus 3	Virus	24	Total
SRP103819	HNEpC	Human Rhinovirus	Virus	48	Total
SRP161185	ATII	Influenza A Virus	Virus	24	Total
SRP230823	HeLa	Influenza A Virus	Virus	24	Total
SRP234025	A549	Influenza A Virus	Virus	48	Total
SRP253951	A549	Influenza A Virus	Virus	9	Total
SRP272285	A549	Influenza A Virus	Virus	6	Total
SRP277269	293T	Influenza A Virus	Virus	6	Total
SRP281173	A549	Influenza A Virus	Virus	12	Total
SRP170549	Calu3	MERS-CoV	Virus	24	Total
SRP227272	Calu3	MERS-CoV	Virus	24	mRNA
SRP096169	HFF	Orf Virus	Virus	8	Total
SRP277439	HEK293	Porcine Rubulavirus	Virus	12	Total
SRP229586	A549	Respiratory Syncytial Virus	Virus	36	Total
SRP229586	H292	Respiratory Syncytial Virus	Virus	36	Total
SRP229586	HBEC	Respiratory Syncytial Virus	Virus	36	Total
SRP253951	A549	Respiratory Syncytial Virus	Virus	24	Total
SRP115192	HSAEpC	Rift Valley Fever Virus	Virus	18	Total
SRP094462	HInEpC	Rotavirus	Virus	6	Total
SRP253951	A549-ACE2	SARS-CoV-2	Virus	24	Total
SRP270817	PHAE	SARS-CoV-2	Virus	48	Total
SRP273473	DC	SARS-CoV-2	Virus	2	Total
SRP273473	MAC	SARS-CoV-2	Virus	2	Total
SRP278618	iPSC-derived cardiomyocyte	SARS-CoV-2	Virus	48	Total
SRP081284	MeWo	Varicella-zoster Virus	Virus	24	Total
SRP225661	A549	West Nile Virus	Virus	24	Total
SRP142592	hNSC	Zika Virus	Virus	72	Total
SRP251704	A549	Zika Virus	Virus	48	Total
SRP253197	HepG2	Zika Virus	Virus	48	Total
SRP296743	PBMC	*Aspergillus fumigatus*	Fungus	24	Total
SRP296743	PBMC	*Candida albicans*	Fungus	24	Total
SRP296743	PBMC	*Rhizopus oryzae*	Fungus	24	Total
SRP285913	HeLa	*Chlamydia trachomatis*	Bacteria	44	Total
SRP321546	DLD-1	*Fusobacterium nucleatum*	Bacteria	24	Total
SRP321940	Primary human trophoblasts	*Listeria monocytogenes*	Bacteria	5	Total
ERP020415	THP-1	*Mycobacterium tuberculosis*	Bacteria	48	Total
ERP115551	hBMECs	*Neisseria meningitidis*	Bacteria	6	mRNA
SRP263458	HUVEC	*Staphylococcus aureus*	Bacteria	16	Total
SRP072326	A549	*Streptococcus pneumoniae*	Bacteria	2	Total

Table 4.3: **The 69 universal response genes in humans**

RefSeq Accession	Gene Symbol
NM_001547	IFIT2
NM_022168	IFIH1
NM_016323	HERC5
NM_014314	DDX58
NM_080657	RSAD2
NM_021127	PMAIP1
NM_001964	EGR1
NM_001945	HBEGF
NM_005532	IFI27
NM_000584	CXCL8
NM_005252	FOS
NM_014330	PPP1R15A
NM_017414	USP18
NM_152542	PPM1K
NM_014470	RND1
NM_006187	OAS3
NM_005101	ISG15
NM_001570	IRAK2
NM_001565	CXCL10
NM_022750	PARP12
NM_020529	NFKBIA
NM_002463	MX2
NM_006820	IFI44L
NM_001561	TNFRSF9
NM_006734	HIVEP2
NM_012420	IFIT5
NM_024119	DHX58
NM_021035	ZNFX1
NM_002228	JUN
NM_017554	PARP14
NM_001432	EREG
NM_012118	NOCT
NM_003764	STX11
NM_002535	OAS2
NM_003733	OASL
NM_003407	ZFP36
NM_007315	STAT1
NM_022147	RTP4
NM_004419	DUSP5
NM_017631	DDX60
NM_000958	PTGER4
NM_004420	DUSP8
NM_016584	IL23A
NM_000201	ICAM1
NM_172140	IFNL1
NM_030641	APOL6
NM_002053	GBP1
NM_052941	GBP4
NM_002462	MX1
NM_138287	DTX3L
NM_015907	LAP3
NM_005514	HLA-B
NM_017633	TENT5A
NM_003641	IFITM1
NM_001165	BIRC3
NM_002999	SDC4
NM_002038	IFI6
NM_004417	DUSP1
NM_001549	IFIT3
NM_006435	IFITM2
NM_006084	IRF9
NM_004335	BST2
NM_006509	RELB
NM_080745	TRIM69
NM_033390	ZC3H12C
NM_003141	TRIM21
NM_002176	IFNB1
NM_003745	SOCS1
NM_006417	IFI44

Table 4.4: **Multiplex TaqMan RT-qPCR assay for monitoring host immune gene signature expression**

Group	Gene Target	Primer Name	Primer sequence (5'->3')	Probe Sequence (5'->3')	Probe Dye
1	CALR	CALR_F	GAGTATTCTCCCGATCCCAGTATCTATGCC	ATGAGGCATACGCTGAGGAGTTTGG	ABY
		CALR_R	ATTTGTTTCTCTGCTGCCTTTGTTACGCCC		
	RACK1	RACK1_F	TCCCACTTTGTTAGTGATGTGGTTATCTCC	CAGTTTGCCCTCTCAGGCTCCT	VIC
		RACK1_R	CAAATCGCCTCGTGGTGGTGCCCGTTGTGAG		
	RNaseP	RNaseP_F	AGATTTGGACCTGCGAGCG	TTCTGACCTGAAGGCTCTGCGCG	FAM
		RNaseP_R	GAGCGGCTGTCTCCACAAGT		
2	DDX58	DDX58_F	CCGGAAGACCCTGGACCCTA	TTAGGGAGGAAGAGGTGCAG	ABY
		DDX58_R	AGGGCATCCAAAAAGCCACG		
	IFIT2	IFIT2_F	CCCTGCCGAACAGCTGAGAA	CTGCAACCATGAGTGAGAAC	VIC
		IFIT2_R	AGTTGCCGTAGGCTGCTCTC		
	IFITM2	IFITM2_F	ATAGCATTCGCGTACTCCGT	TGCCTCCACCGCCAAGTGC	FAM
		IFITM2_R	TGATGCCTCCTGATCTATCGC		
3	Mx1	Mx1_F	TAGAGAGCTGCCAGGCTTTG	TACACACCGTGACGGATATG	ABY
		Mx1_R	ATCTGTGAAAGCAAGCCGGA		
	IFI6	IFI6_F	TCGCTGCTGTGCCCATCTATC	CTGCTGCTCTTCACTTGC	VIC
		IFI6_R	TTCTTACCTGCCTCCACCCCAC		
	IFIT3	IFIT3_F	ACAGCAGAGACACAGAGGGCA	TCATGAGTGAGGTCACCAAG	FAM
		IFIT3_R	AGCTGTGGAAGGATTTTCTCCAGG		
4	IFI27	IFI27_F	GCCACGGAATTAACCCGAGC	CATCAGCAGTGACCAGTGTG	ABY
		IFI27_R	GCCACAACTCCTCCAATCACA		
	IFIH1	IFIH1_F	ACAGCTTCACCTGGTGTTGGA	CGAAGCAAGCCAAAGCTGAAG	VIC
		IFIH1_R	ATGGCAAACTTCTTGCATGGCT		
	PARP12	PARP12_F	ACCATGCAAACCTGCAATACC	TCCAGGCCCGAAGAGCATC	FAM
		PARP12_R	GCAGCGTGCGGTTAAAGAG		
5	IRF9	IRF9_F	GCTCTTCAGAACCGCCTACTTC	CTCCAGCCATACTCCACAGAATC	ABY
		IRF9_R	CTCCAGCAAGTATCGGGCAA		
	CXCL10	CXCL10_F	TGCAAGCCAATTTTGTCCACG	AGCAGTTAGCAAGGAAAGGTC	VIC
		CXCL10_R	GCCTCTGTGTGGTCCATCCT		
	Mx2	Mx2_F	CATGATTGTGAAGTGCCGGG	CTGAGCTTGGCAGAGGCAAC	FAM
		Mx2_R	CAACGGGAGCGATTTTTGGA		
6	OAS2	OAS2_F	CGTTGGTGTTGGCATCTTCTG	CCAGTCCCATCCTTGAAGCAG	ABY
		OAS2_R	TGCATTGTCGGCACTTTCC		
	CXCL8	CXCL8_F	CCAGGAAGAAACCACCGGAA	TGGCCGTGGCTCTCTTG	VIC
		CXCL8_R	CTTGGCAAAACTGCACCTTCAC		
	RTP4	RTP4_F	TGGACGCTGAAGTTGGATGGC	CTCTCTGTTGGTATTGCTTC	FAM
		RTP4_R	CAACTTCGCTGGCAGGAGGAA		

Table 4.5: **Top 30 differentially up- and down- regulated genes from comparison between infected and healthy saliva**

Gene Symbols	Log2(Fold Change)	Adjusted P-value
CHRNA5	6.05	9.35E-76
IL2RA	6.07	1.08E-71
STS	6.02	7.91E-69
BAG5	5.80	9.31E-64
HBD	7.01	3.53E-53
POR	6.03	4.83E-50
LCN10	6.38	4.06E-46
C10orf55	7.06	9.76E-44
TWIST1	6.35	1.08E-43
CA2	6.97	1.19E-43
NR0B1	7.13	7.96E-43
GALE	5.83	1.04E-42
TENT5A	6.15	2.69E-42
WRN	5.11	3.91E-42
NOS3	5.95	5.09E-41
HBEGF	5.00	8.94E-41
DRD4	6.13	5.62E-40
NCMAP	6.31	3.29E-39
REN	5.61	7.10E-39
FGG	4.98	2.07E-37
HADHA	5.01	8.57E-37
HBG2	7.61	2.11E-36
HOXD13	4.86	2.50E-36
KITLG	5.31	1.18E-35
CHRNB1	5.74	1.08E-32
ITGB3	4.59	2.63E-32
BST2	6.03	3.66E-32
OR56B1	7.34	4.66E-31
HBG1	8.01	5.45E-31
RND1	7.31	6.27E-31

Gene Symbols	Log2(Fold Change)	Adjusted P-value
LOC102723665	-3.38	1.86E-06
GCSAM	-4.12	1.84E-05
TAAR9	-5.50	2.94E-05
CDCA7L	-3.59	1.16E-04
MIR320B2	-4.81	1.47E-04
HULC	-5.84	1.49E-04
ZNF235	-3.25	2.40E-04
SLC39A12	-3.05	3.28E-04
IVNS1ABP	-3.87	3.58E-04
KLHDC4	-3.96	4.01E-04
SERPINB5	-3.57	4.41E-04
LOC101927143	-4.42	4.45E-04
VAV2	-3.29	4.68E-04
DSEL	-4.39	5.69E-04
RPL22	-2.67	7.18E-04
LINC01085	-3.48	7.23E-04
ERVW-1	-3.94	8.02E-04
SLC25A25-AS1	-3.54	8.58E-04
THOC5	-2.59	9.56E-04
UXT-AS1	-4.49	1.21E-03
TRI-AAT1-1	-3.34	1.37E-03
AKAP4	-3.07	1.76E-03
TADA2A	-2.58	2.03E-03
LRRC7	-3.49	2.71E-03
LEMD1-AS1	-3.55	3.02E-03
GNG14	-3.82	3.37E-03
ZNF461	-3.55	3.77E-03
LINC01781	-2.66	4.07E-03
SAMD13	-3.46	4.65E-03
SLAMF8	-1.81	5.00E-03

Chapter 5

Macrophages derived from human induced pluripotent stem cells serve as a relevant and tractable model for viral-host interaction

---

Adapted from: Yang Q, Warren CJ, Fairchild LM, Rowland TJ, Nahreini TS, Allen MA, Dowell RD, Sawyer SL. Macrophages derived from human induced pluripotent stem cells serve as a relevant and tractable model for viral-host interaction. In Preparation. 2022.

5.1 Contributions

This project was supported by the University of Colorado Shared Resource Grant in an effort to encourage the collaboration with the shared core facilities around the campus. Dr. Cody Warren came up with this idea of validating stem cell derived monocytes and designed the framework of the project. Indeed, the project would not be possible without the support from several shared core facilities, including the stem cell core, next gene sequencing core and flow cytometry core. And the key drivers are the wonderful scientists who work at the core facilities. Specifically, Lawrence Fairchild and Dr. Teisha Rowland were the pioneers who established the monocyte differentiation protocol before passing it to me. Theresa Nahrenini supported us in all flow cytometry analyses, especially in the early days of isolating monocytes from blood. During the project, I carried out the stem cell differentiation to generate the monocytes that were eventually used for this study. I also work alongside with Dr. Cody Warren on obtaining blood and blood-derived monocytes. To characterize this

unique cell type, I cultured the cells, prepared the RNA and ATAC sequencing libraries, and analyzed the sequencing data. To finally validate the model system in the context of viral infections, I worked alongside with Dr. Cody Warren to carry out the HIV-1, Dengue virus and influenza virus infections. The manuscript that follows is the collective result of tireless contribution by all listed co-authors.

5.2 Abstract

Monocytes are widespread in the bloodstream and are often the first population of immune cells that contact viruses. Further, as a mechanism of host defense, monocytes differentiate into long-lived macrophages that function as key mediators of inflammation and antiviral immunity. However, this cell lineage sometimes act as a double-edged sword during viral infections; monocytes and macrophages can be manipulated by viruses and serve as vessels for viral replication, dissemination, and long-term persistence. Thus, the study of monocyte/macrophage-virus interactions is essential for our understanding of human diseases. However, these cells have relatively low abundance in the bloodstream and cannot be propagated ex vivo, meaning that consistent access to blood donors is required. To overcome these hurdles, we evaluated monocyte-derived macrophages from induced pluripotent stem cells (iPSCs) as an alternative model system. Since iPSCs proliferate indefinitely and can be expanded to large quantities, we were able to isolate >40 million iPSC-derived monocytes over a short period, a feat that would have otherwise required liters of blood from numerous donors. We show that, through phenotypic profiling of surface marker expression, and transcriptomic (RNA-seq) and genome accessibility (ATAC-seq) profiling, iPSC-derived macrophages are nearly identical to their blood-derived counterparts. Furthermore, we show that iPSC-derived macrophages support the replication of viruses including HIV-1, dengue virus, and influenza virus. Our results implicate iPSC-derived macrophages as a valuable new model system for studying host-virus interactions.

5.3 Introduction

Cell culture models for the study of viruses often depend on immortalized or transformed (tumor-derived) cell lines. These cells are beneficial for experimentation because they tend to divide quickly, are inexpensive to maintain, and proliferate indefinitely. However, they oftentimes fail to appropriately recapitulate the biology of cells derived directly from a host. This is because many immortalized/ transformed cells have accumulated mutations in genes that control important facets of cell biology, including antiviral defense [192]. Defects in cellular antiviral defense mechanisms will have impacts on viral replication that may not be physiologically relevant. Thus, an ideal *in vitro* cell culture model system for studying virus biology should be representative of cells *in vivo*.

Primary cell culture models best represent cells of their original tissue. Following viral exposure, various myeloid cell subsets are known to play central roles in the antiviral immune response and viral clearance. Monocytes are one such subset. Monocytes are widespread in the bloodstream and are often the first population of immune cells that contact viruses. Further, as a mechanism of host defense, monocytes differentiate into long-lived macrophages that function as key mediators of inflammation and antiviral immunity. Many viruses, in turn, hijack monocytes and macrophages and exploit them as vessels for viral replication and dissemination throughout the body [20]. Thus, these cells sometimes act as a double-edged sword during viral infections; monocytes and macrophages are important for antiviral immunity, but can also be manipulated by viruses to ensure their propagation and transmission. Studying the biology of viruses in natural host cell targets like monocytes and macrophages is essential to provide physiologically relevant insights into virus biology and virus-driven diseases.

Primary monocytes and macrophages are not a particularly tractable model system. Blood circulating monocytes are non-proliferative [193], are in low abundance (3-8% of blood mononuclear cells) [194], and standard isolation procedures vary in purity and yield, thereby

introducing variability from one experiment to the next. In addition, acquiring these cells requires consistent access to blood donors or their purchase from commercial sources, both of which come at considerable cost. Cells differentiated from induced pluripotent stem cells (iPSCs), however, overcome many of these limitations. iPSCs are a type of pluripotent stem cell—master regulator cells capable of giving rise to different cell types—that are obtained from somatic cells through genetic reprogramming [195, 196]. Since iPSCs proliferate indefinitely, large numbers of cells can be obtained from a single clone, providing much needed consistency between experiments. iPSCs can also be differentiated into a variety of cell types that resemble primary cells, enabling much needed biological relevance. Finally, iPSCs can be genetically modified with tools like CRISPR, and the engineered traits subsequently passed on to a desired differentiated cell type [197, 198], thus enabling genetic modification of physiologically relevant cell types that would be otherwise extremely difficult to manipulate *in vitro*. When taken together, generating monocytes and monocyte-derived macrophages from iPSCs has the potential to overcome many of the difficulties associated with primary blood-derived cells.

Here, we evaluate a kit-based method for producing monocytes from iPSCs. Following iPSC differentiation, we harvest CD14-positive monocytes and then differentiate them into macrophages. We then compared blood- and iPSC-derived cells using several metrics, including surface marker expression by flow cytometry, global transcriptome profiling (RNAseq), and genome-wide chromatin accessibility (ATACseq). Our results confirm that these iPSC-derived cells are virtually indistinguishable from their blood-derived counterparts. To assess their utility as virus infection models, we challenged iPSC-derived macrophages with a panel of clinically relevant RNA viruses (dengue virus, HIV-1, and influenza virus) and found that iPSC-derived macrophages faithfully recapitulate the viral replication kinetics observed in blood monocyte-derived macrophages. These methods should prove invaluable in answering questions relating to human monocyte and macrophages roles in virology, given their ease of access, reproducibility, and biological relevance.

5.4 Results

5.4.1 Monocytes and macrophages differentiated from iPSCs are phenotypically similar to primary blood-derived cells.

Myeloid cells, including monocytes and macrophages, are differentiated from hematopoietic stem progenitor cells (HSPCs). Here, we used a commercially available differentiation kit to generate these HSPCs from induced pluripotent stem cells (iPSCs). Following the manufacturers directions, these progenitor cells were then differentiated into monocytes, which were harvested directly from the cell culture supernatant. To establish the robustness of this method, we evaluated monocyte yields from two patient-derived iPSC cell lines over a 4-week timecourse, with serial harvests every 4-5 days. The total average number of monocytes produced from four wells of a 6-well dish was 40.1×10^6 cells (n = 2 iPSC lines iCTR and iC7-2). After an initial burst of monocytes in the first two weeks (28.2×10^6 total cells), we observed a gradual decline in cell numbers over the remaining two weeks (7.5×10^6 cells in week 3 and 4.4×10^6 cells in week 4). In contrast, single 50 ml blood harvest yielded 2.0×10^6 of monocytes on average (n=6 harvests).

To determine if iPSC-derived monocytes were similar to their blood-derived counterparts, we assessed cell morphology by bright-field microscopy and surface marker expression by flow cytometry. Monocytes derived from both iPSCs and blood were adherent, large (12-21 μm in diameter), and round to oval shaped (Fig. 5.1A). Blood- and iPSC-derived monocytes expressed canonical surface markers CD14 and CD11b (Fig. 5.1B). Although the purity of iPSC-derived monocytes was lower than blood-derived cells (Figure 5.2A, average CD14+ cells 75.2%, n = 2 iPSC lines vs. 93.1% from blood), our results were consistent with the manufacturers expectations. We next determined the capacity for iPSC-derived monocytes to differentiate into macrophages. Adhered monocytes were first treated with macrophage colony stimulating factor (M-CSF) for four days to generate naiive (M0) macrophages. M0 macrophages were then polarized into M1 (pro-inflammatory) or M2 (anti-inflammatory)

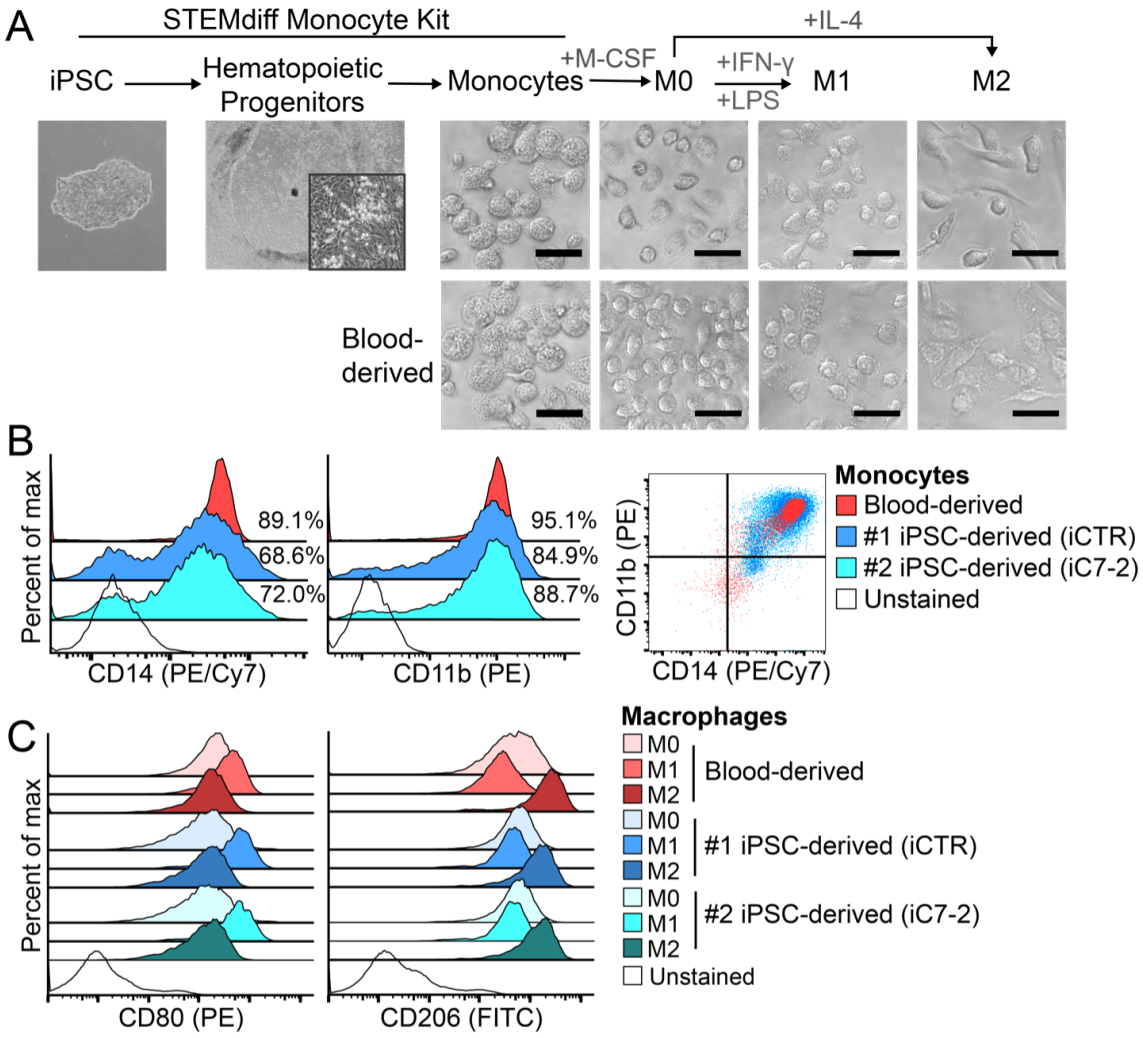


Figure 5.1: **Monocytes and macrophages differentiated from iPSCs are phenotypically similar to primary blood-derived cells. (A)** iPSCs were differentiated into monocytes following the manufacturer's instructions (STEMdiff monocyte kit, see methods). Blood was drawn from healthy donors and monocytes were isolated through a series of density gradient purifications (see methods). Subsequently, both blood- and iPSC-derived monocytes were differentiated into naïve M0 macrophages via 10 ng/mL M-CSF treatment for 4 days. Naïve macrophages were then polarized into M1 macrophages using 50 ng/mL IFN-γ and 10 ng/mL LPS, or into M2 macrophages using 10 ng/mL IL-4 for 48 hours. The morphology comparison between different monocytes and macrophage subtypes are shown as bright field images. Scale bars equal 40 μm. **(B, C)** The expression levels of surface proteins were quantified via fluorescent antibody staining followed by flow cytometry (1 representative experiment shown from 3 biological replicates). Monocytes and the subsequent macrophage subtypes derived from blood and two unique patient-derived iPSC lines were used for the analysis. **(B)** The expression levels of monocyte markers CD14 and CD11b were quantified for freshly isolated monocytes, and following macrophage polarization **(C)** the expression levels for CD80 (M1 macrophage marker) and CD206 (M2 macrophage marker) were quantified within the CD14+ macrophage subset.

subsets using IFN-γ and LPS (lipopolysaccharide) or IL-4, respectively. iPSC monocyte-derived macrophages were similar in morphology to primary monocyte-derived macrophages (Fig. 5.1A), and expressed similar levels of CD80 (M1 subset marker) and CD206 (M2 subset marker) surface markers (Fig. 5.1C, Figure 5.2B). When taken together, iPSC-derived monocytes and monocyte-derived macrophages are similar in morphology and surface marker expression when compared to their primary cell counterparts.

5.4.2 Blood- and iPSC-derived monocytes and macrophages demonstrate similar transcriptomic change during differentiation and polarization

To examine whether the blood- and iPSC-derived monocytes/macrophages exhibit similar transcriptional profiles, we carried out transcriptomic and chromatin accessibility analyses of all cell types during differentiation and subsequent polarization via bulk RNA-seq and ATAC-seq. Through RNA-seq analysis, we observed distinct transcriptomic profiles that are representative of the cell types during differentiation/polarization, as the primary (PC1, 56.9% data variance) and the secondary (PC2, 18.0%) principle components clearly separate the data sets based on the cell type (indicated in colors in Fig. 5.3A). On the other hand, the tissue origins (blood or iPSC, indicated in symbol shapes in Fig. 5.3A) or the biological replicates had minimal effects on the data variance. In addition, via pair-wise transcriptomic comparisons of blood- and iPSC-derived monocytes and macrophages , we observed strong correlation of gene expressions between blood- and iPSC-derived cells: monocytes (Pearson Correlation Coefficient, PCC=0.95), M0 (PCC=0.95), M1 (PCC=0.95) and M2 (PCC=0.94) (Fig. 5.3B). Together, this suggests that blood- and iPSC-derived monocytes/macrophages have similar transcriptomic profiles at each differentiation/polarization stage, and that they are equally capable of being polarized into the desired macrophage subtypes, independent of their tissue origins.

Furthermore, comparisons across distinct cell types yielded gene sets undergoing significant transcriptional changes. We closely examined the genes that are differentially ex-

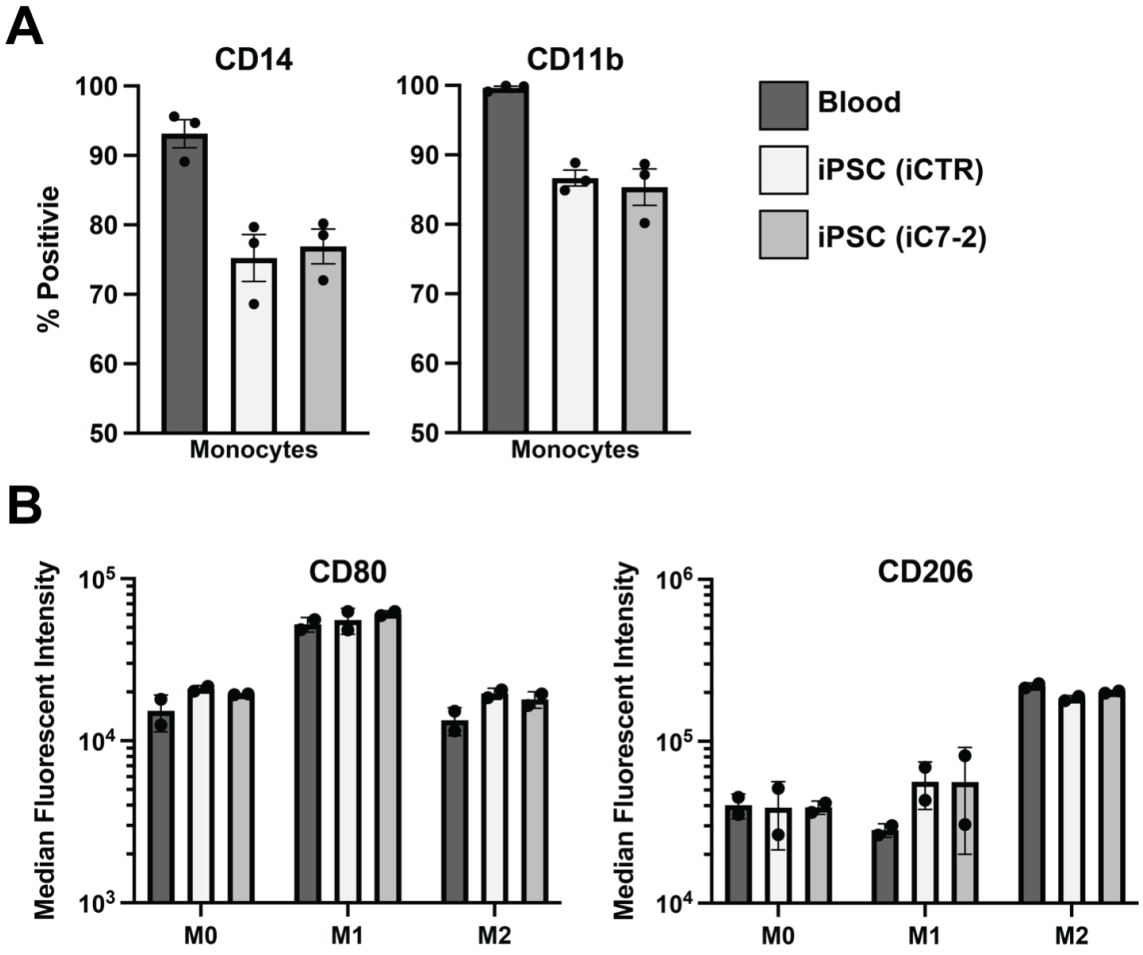


Figure 5.2: **Comparison of surface marker expression between blood- and iPSC-derived monocytes.** iPSCs were differentiated into monocytes following the manufacturer's instructions (STEMdiff monocyte kit, see methods). Blood was drawn from healthy donors and monocytes were isolated through a series of density gradient purifications (see methods). The expression levels of surface proteins were quantified via fluorescent antibody staining followed by flow cytometry (error bar indicates SEM for 2-3 biological replicates). **(A)**The expression levels of monocyte markers CD14 and CD11b were quantified for freshly isolated monocytes. **(B)** The expression levels of CD80 and CD206 were quantified using polarized macrophage subtypes derived from blood or IPSC. n.s.: not significant, **: p-value$\leq$0.01 via two-tailed Student's t-test between Blood- and iPSC- derived macrophage.

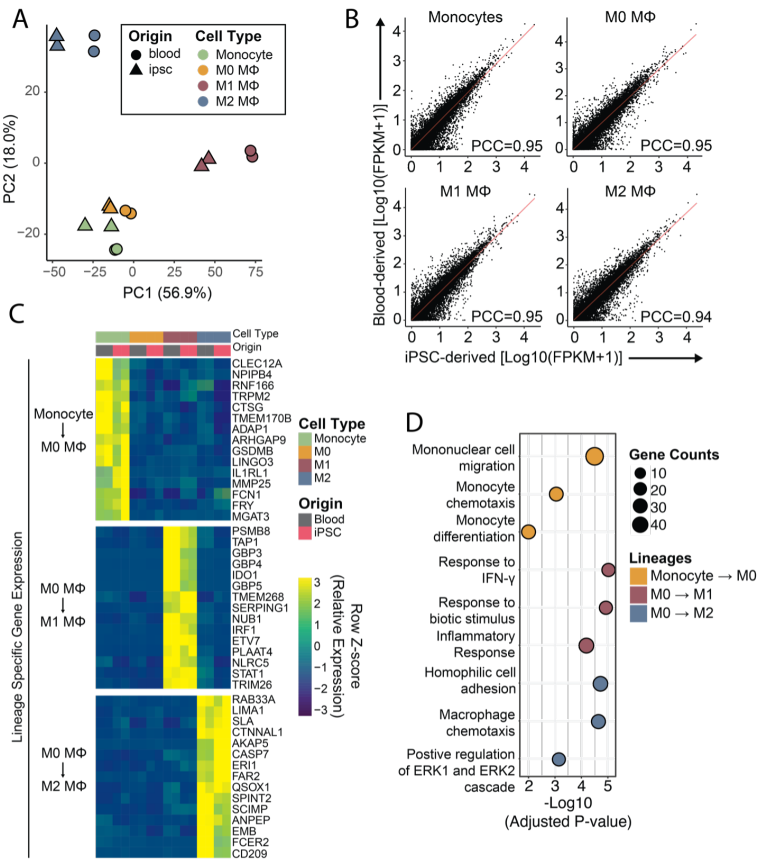


Figure 5.3: **Blood- and iPSC-derived monocytes and macrophages share similar transcriptomic profiles during differentiation. (A)** Principle component analysis (PCA) of the transcriptomic profiles of blood- and iPSC-derived monocyte/macrophage cell types. The first two principle components are shown with the percentages of variance indicated in parentheses. The specific cell types are indicated by different colors, while the origin of the cell types are indicated by shapes (green, monocytes; orange, M0 macrophages, maroon, M1 macrophages; blue, M2 macrophages. Triangle, iPSC derived; circle, blood derived). **(B)** Scatter plots of the log10-transformed fragments per kilobase of transcript per million (FPKM) values for all RNA transcripts from each cell type that was either derived from blood (y-axis) or iPSC (x-axis). The red line is the linear regression curve of the x and y values, and the Pearson correlation coefficients of the linear regression is indicated on each plot. **(C)** Heatmap of the top 15 differentially expressed genes from each pair-wise comparison representing the genes undergoing significant changes during monocyte differentiation or macrophage polarization. The library size and FPKM-normalized expression data is further scaled to row mean. Each row represents an individual gene, and each column represents the average expression of two biological replicates, where the cell type and origin are color-coded at the top. **(D)** Gene ontology (GO) enrichment analysis of genes undergoing significant expression change during monocyte differentiation or macrophage polarization. The top three enriched GO terms are indicated on each row. The data points shows the degree of enrichment, where the size of the symbol indicate number of genes included, and the x-axis indicates -log10 transformed adjust p-values from the enrichment analysis.

pressed during monocyte to macrophage differentiation (monocytes vs. M0 macrophages) and macrophage polarization (M0 vs. M1 macrophages, M0 vs. M2 macrophages) through pair-wise comparisons (Fig. 5.3C). These transcriptional changes are consistent with previously described, cell-type-specific gene signatures, regardless of iPSC or blood origin. First, comparing the monocytes to naïve (M0) macrophages resulted from M-CSF treatment, we noticed a group of genes that defines classical monocytes undergoing down regulation (CTSG, FCN1, IL1RL1). Gene ontology enrichment analysis indicated that the down-regulated genes from monocytes to M0 macrophages are related to monocyte chemotaxis and differentiation (Fig. 5.3D) [199–201]. As expected, when naïve macrophages are polarized towards the inflammatory M1 macrophage subtype via pro-inflammatory treatment of LPS and IFN-γ, we observed consistent upregulation of genes involved in inflammatory responses (IRF1, IL6, TNF etc.) as well as interferon-gamma stimulated immune genes (GBP5, OAS1, STAT1 etc.) (Fig. 5.3C & D) [17, 202–204]. In the case of M0 to M2 macrophage polarization, the classical IL-4 stimulated genes (TIMP3, CCL24, and CASP7) were similarly upregulated in both blood- and iPSC-derived M2 macrophages [205–207]. Collectively, the comparative RNA-seq analysis shows iPSC- and blood-derived monocytes and macrophages display highly similar cell-type defining gene expression profiles.

5.4.3 Blood- and iPSC-derived macrophages demonstrate similar chromatin accessibility changes during differentiation

To determine whether blood and iPSC-derived monocytes/macrophages exhibit similar genome-wide chromatin accessibility landscapes, and hence to infer the essential transcription factors that are essential for monocyte differentiation and macrophage polarization, we carried out an Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq). By first identifying open chromatin regions, then comparing the normalized read counts within each open chromatin region, we found similar chromatin accessibility profiles among the blood- and iPSC-derived monocytes and macrophage subtypes (Fig. 5.4A). We carried

out correlation analyses of the scatters with a focus on open chromatin regions with read coverage (FPKM) greater than 10. From this analysis, we observed a strong correlation of chromatin accessibilities between blood- and iPSC-derived monocytes (PCC=0.91), M0 (PCC=0.89), M1 (PCC =0.82) and M2 (PCC =0.63) macrophages. The relatively lower correlation between blood and iPSC-derived M2 macrophages is likely the result of heterogeneity during macrophage polarization [17]. Furthermore, to unbiasedly characterize the accessibility change within these open chromatin regions across the macrophage lineages, we carried out hierarchical clustering of the accessibility profiles among the relevant cell types to parse out patterns of chromatin accessibility changes. Consistent with the correlation analysis, blood- and iPSC-derived monocytes and macrophages have similar chromatin accessibility profiles within the same cell types, but differ distinctively across cell types (Fig. 5.4B, "AC" denotes ATAC-seq clusters). For example, within AC1 cluster, we identified a group of chromatin regions that are only highly accessible in iPSC progenitors but remained relatively closed across the monocyte/macrophage lineages, and vice versa in the AC3 cluster. We also identified accessible chromatin regions that are specific to monocytes (AC4), and M1 (AC6) and M2 (AC7) macrophages. Interestingly, we also found a group of chromatin regions that are more accessible in iPSC and M2 macrophages but not in other cell types (Figure 5.4B, AC2). Together, through unbiased clustering, we identified unique ATAC-seq open chromatin regions representative of each cell type. These cell-type-defining open chromatin regions are conserved from blood to iPSC derived cell types.

Looking more closely, when combining the ATAC-seq and RNA-seq datasets together, we observed consistency between open chromatin regions and cell type-specific expression of key gene signatures for specific cell types (Fig. 5.4C). For example, Ficolin-1 (FCN1) is known to be highly expressed in blood circulating classical monocytes [208]. We found that the expression of FCN1 distinguishes monocytes from monocyte-derived macrophages in both blood- and iPSC-derived systems (left in Fig. 5.4C). This is likely the result of transcription factor binding activities happening upstream of the FCN1 gene. Consistent with this asser-

tion, we observed similar chromatin footprints upstream of the FCN1 gene locus between blood- and iPSC-derived monocytes that were absent in macrophages. These observations were further extended to M1 macrophage-specific IRF1 transcription and M2 macrophage-specific GATA3 transcription (center and right in Fig. 5.4C, respectively) [209–211]. Together, this suggests not only do blood- and iPSC-derived monocytes and macrophages feature similar transcriptomic profiles, this similarity is also the result of highly similar transcription regulatory network.

Finally, to determine the underlying transcription factors that are likely dictating the unique transcription profiles among blood- and iPSC-derived monocyte/macrophage lineages, we carried out motif enrichment analysis on identified open chromatin region clusters (Fig. 5.4D). Consistent with the previous understanding, we observed strong SOX2 and NANOG motif enrichment in AC1, the open chromatin cluster specific to iPSC progenitors [37]; PPARγ and STAT6 enrichment for cluster AC3, representing chromatin regions that remained accessible across myeloid lineages [212]; enrichment of STAT1 and IRF3 motifs in M1 macrophage-specific open chromatin regions (AC6) [205]; and the enrichment of STAT6, RUNX1/2, CEBPα/ β, and PU.1/SPI1 family motifs in chromatin regions unique to M2 macrophages (AC2 and AC7) [213]. Together, this highlights the similarity of transcriptional networks underlying the blood- and iPSC-derived monocyte/macrophages cell types.

5.4.4 Blood- and iPSC-derived macrophages support persistent HIV-1 infection

Now that we have demonstrated that iPSC- and blood-derived macrophages are phenotypically similar and nearly identical in their transcriptomic and chromatin accessibility profiles, we sought to assess their utility as model systems for diverse viral infections. First, we started with HIV-1. HIV-1 tropism is usually restricted to T cells that express the required receptor (CD4) and co-receptors (CCR5 or CXCR4) used for virus entry. However, during the course of HIV-1 infection, viral variants emerge that have the capacity to infect

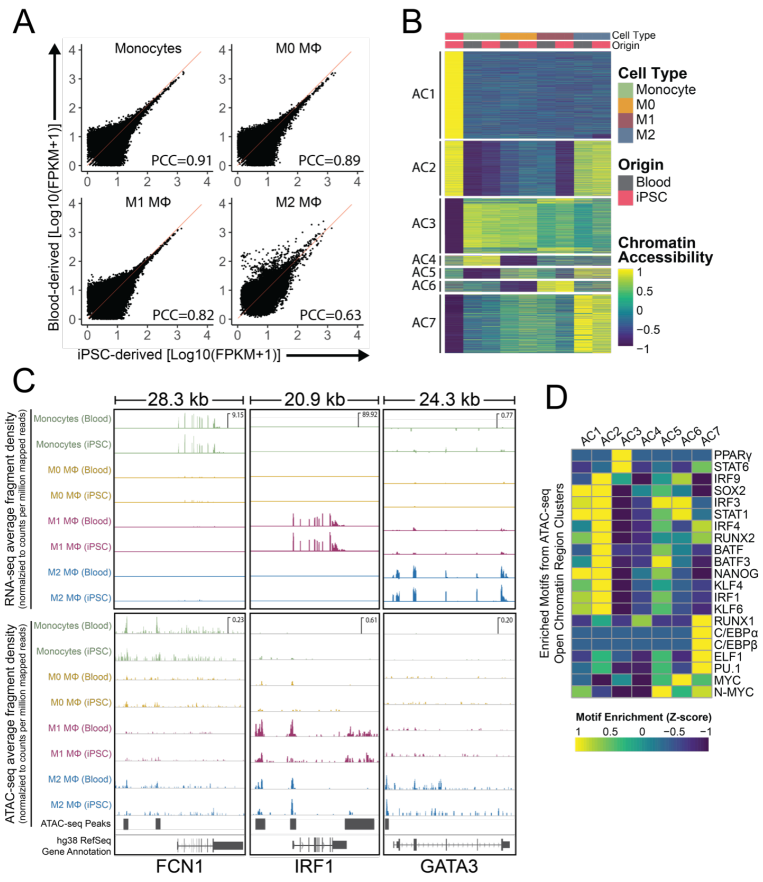


Figure 5.4: **Blood- and iPSC-derived monocytes and macrophages share similar chromatin accessibility profiles following differentiation and polarization.** **(A)** Scatter plots of the log10-transformed fragments per kilobase of transcript per million (FPKM) values for all ATAC-seq peaks from each cell type that was either derived from blood (y-axis) or iPSC (x-axis). **(B)** Heatmap of open chromatin regions that demonstrated significant accessibility changes during monocyte differentiation or macrophage polarization. The regions are assigned to clusters AC1-AC7 through hierarchical clustering (shown on the left). The library size and FPKM-normalized read counts for each ATAC-seq peak is further scaled to the row mean. Each row represents an ATAC-seq peak, and each column is the average chromatin accessibility of two biological replicates, where the cell type and origin are color-coded at the top. **(C)** Integrated Genome Viewer (IGV) screenshots of chromatin accessibility changes along with gene expression changes of representative genes (FCN1, IRF1 and GATA3) during monocyte differentiation and macrophage polarization. Each track is a bargraph representing the counts-per-million-mapped-reads (CPM)-normalized read coverage over ATAC-seq peaks or annotated genes. Each track is also representative of the two biological replicates. The track heights are group auto-scaled, and the scale for each group is indicated on the first track. **(D)** Transcription factor motif enrichment analysis of each open chromatin region clusters (shown in B, represented here in each column) that demonstrates significant accessibility changes during monocyte differentiation or macrophage polarization. The enriched transcription factor motifs that are relevant to macrophage lineages are represented on each row, and annotated on the right. The adjusted p-value from the motif enrichment is -log10 transformed and then row mean-normalized.

monocyte-derived and tissue-resident macrophages. These cells are thought to be important, long-term viral reservoirs *in vivo* (reviewed in [214]), and thus exploring HIV-1 biology in these clinically relevant cells would provide valuable insights into the role macrophages play as reservoirs for HIV-1 infection.

To comparatively assess iPSC- and blood-derived macrophages for their susceptibility to HIV-1 infection, we utilized unpolarized (M0) macrophages as our model. Unpolarized macrophages are generally susceptible to HIV-1 infection, whereas M1- or M2-polarized macrophages are refractory [34]. We infected these cells with HIV-1 isolate SF162, a prototypical macrophage-tropic HIV. Seven days post infection, the cell supernatant was used to infect TZM-bl indicator cells (a CD4/CCR5 expressing cell line that produces luciferase in response to HIV-1 gene transcription, [215]) to monitor for the presence of virus in the cell supernatant. Additionally, HIV-1 exposed macrophages were washed extensively, then lifted from the dish and co-cultured with TZM-bl cells for 3 days, thus enabling cell-cell infection by HIV-1 (Fig. 5.5A). Virus released from macrophages into the culture supernatant was infectious in TZMbl cells, and the magnitude of this infection was comparable between culture supernatant obtained from HIV-1 exposed blood- and iPSC-derived macrophages (Fig. 5.5B). Further, macrophages were demonstrated to be persistently infected, as the macrophages used in co-culture were fully viable and capable of transmitting virus to naïve cells (Fig. 5.5C). When taken together, iPSC-derived macrophages can be substituted for blood-derived cells as a suitable model system to study HIV-1 infection.

5.4.5 Blood- and iPSC-derived macrophages support high titer dengue virus production

Monocytes and macrophages are known to support dengue virus replication. Following the initial exposure to dengue virus, monocyte/macrophage-mediated antibody-dependent infection enhancement is the key that allows the virus to establish the systemic infection and subsequently leads to severe symptoms [23]. We next sought to test the capacity of iPSC-

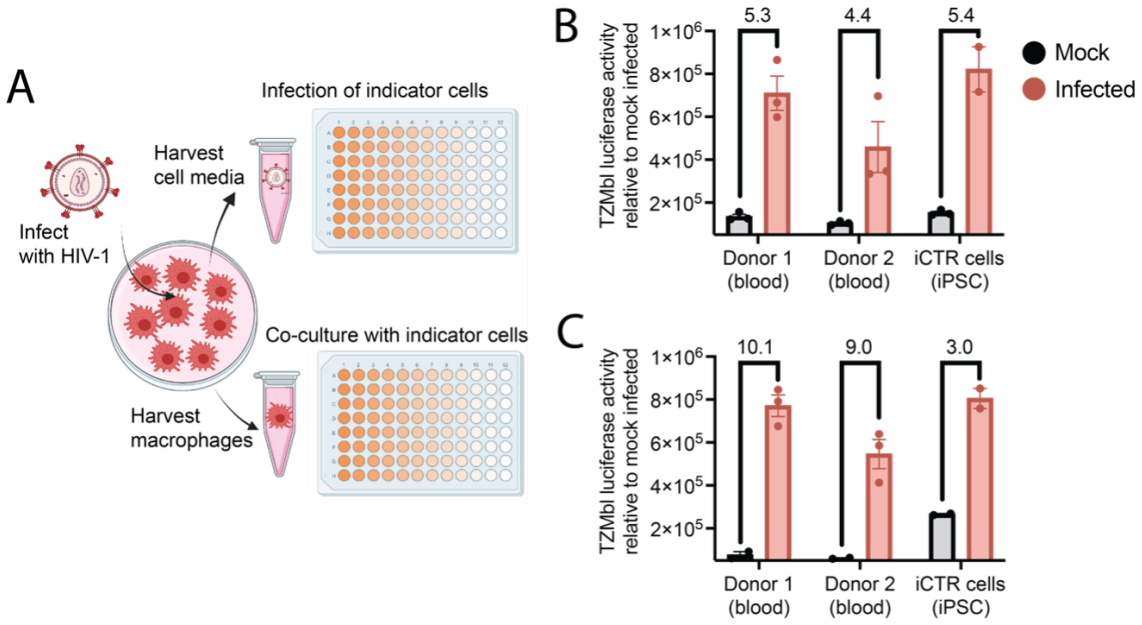


Figure 5.5: **iPSC- and blood-derived macrophages are equally susceptible to HIV-1 infection.** **(A)** Blood- and iPSC-derived monocytes were differentiated into M0 macrophages using M-CSF. Macrophages were infected with HIV-1 isolate SF162 and then seven days later the virus-containing cell supernatant was removed and used to infect TZMbl indicator cells for 2 d **(B)**. In parallel, macrophages were washed extensively, lifted from the culture dish and then co-cultured with TZMbl indicator cells for 3 d **(C)**. TZMbl indicator cells respond to HIV-1 infection by producing luciferase, an enzyme whose activity can be quantitatively assessed by light output (Y-axis values). Error bars represent the mean +/- SD of technical triplicates from 1 (mock) or 2-3 biological replicates (HIV-1 infected cells). Values above bars represent fold change relative to mock infected cells.

derived macrophages to serve as a relevant model for dengue virus infection. We infected blood- and iPSC-derived M0 macrophages with dengue virus 2 (DENV2 Thailand/16681/84) at MOI of 0.1 and enumerated viral titers in the cell supernatant by plaque assay over a time course (Fig. 5.6A). We observed that both blood- and iPSC-derived macrophages supported dengue virus replication to similar endpoint titers and growth kinetics (Fig. 5.6B). We then tested the effects of M1 macrophage polarization on dengue virus replication. LPS and IFN-γ, both used in inducing M0-to-M1 polarization, are known to render cells refractory to dengue virus infection [216–218]. In line with these studies, we observed that M1 macrophages showed significant restriction on dengue virus replication compared to M0 and M2 macrophage subtypes (Fig. 5.6C). We speculate that such differences can be attributed to certain dengue virus restriction factors within the inflammatory macrophage subtypes. Collectively, our results suggest that dengue virus infection is appropriately modeled in iPSC-derived macrophages, further emphasizing the important utility of these cells.

5.4.6 Blood- and iPSC-derived macrophages restrict influenza virus production at a late stage of virus infection

It was previously shown that monocyte-derived macrophages from blood exhibit potent restriction against most seasonal influenza virus strains, although the exact mechanism remains unknown [219]. Here, we sought to test if iPSC-derived macrophages also restrict influenza virus production, and to characterize the state at which virus infection is restricted. First, we infected blood- and iPSC- derived naïve macrophages with influenza virus (A/Udorn/307/1972) at an MOI of 0.1 and 0.01 and monitored for infectious virus production by plaque assay 24 h later (Fig. 5.7A). Notably, we found that influenza virus production is limited in both iPSC- and blood-derived macrophages (Fig. 5.7B), further providing evidence that iPSC-derived macrophages fully recapitulate phenotypes observed in blood-derived cells. Next, we repeated viral growth assays in several control cell lines: 1) THP-1 and U937 are monocytic cell lines that can be differentiated into macrophages using

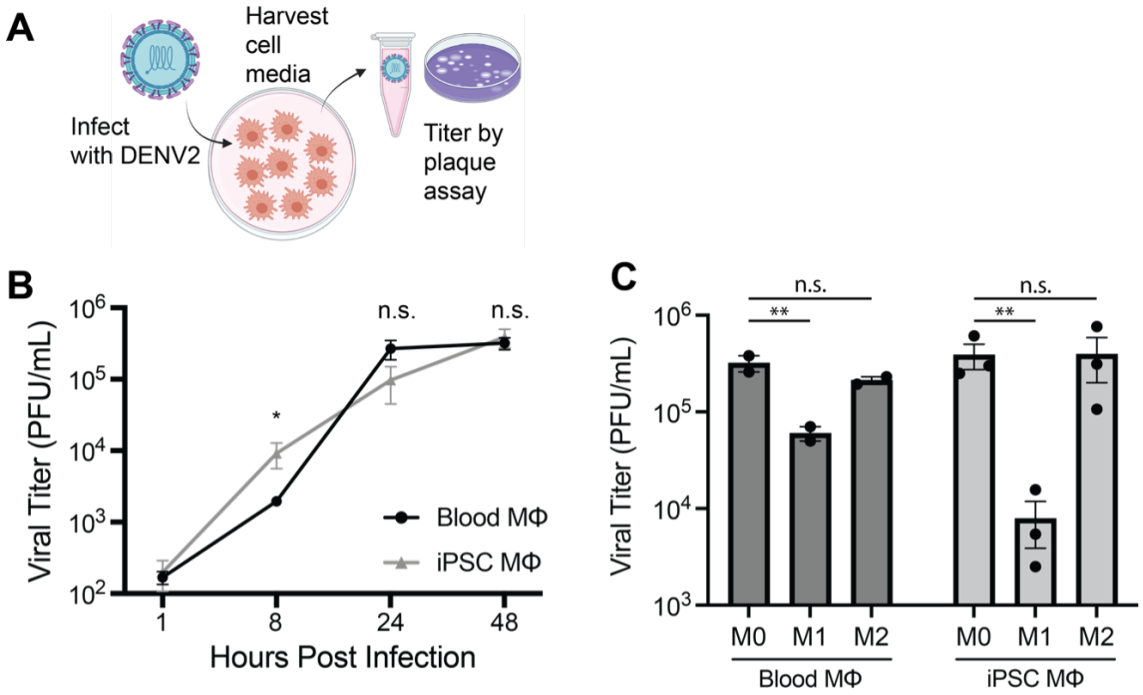


Figure 5.6: **Blood- and iPSC-derived macrophages support high titer dengue virus production. (A, B)** Blood- and iPSC-derived M0 macrophages were infected with dengue virus 2 at an MOI of 0.1 and the cell culture supernatant was harvested and titered by plaque assay at the indicated time points post infection (X-axis). **(C)** Blood- and iPSC-derived macrophages subtypes, M0, M1, and M2, are infected with Dengue virus 2 at MOI of 0.1 for 48 hours, and the supernatant from the infection was harvested and titered by plaque assay. Error bars represent SD of 2-3 biological replicates. n.s.: not significant, *: p-value ≤ 0.05, **: p-value ≤ 0.01 via two-tailed Student's t-test between pair-wise viral titers comparison.

PMA (phorbol 12-myristate-13 -acetate) and 2) A549 cells are common lung epithelial cells that are susceptible to influenza virus infection. We observed significant virus production in both PMA-differentiated macrophages and in A549 cells (Fig. 5.7B). When taken together, our results indicate that influenza virus infection is robustly restricted by a host factor that is specifically expressed in primary and iPSC-derived macrophages, but not in commonly used cell lines for influenza virus propagation.

We then sought to address at which stage of infection influenza virus is being restricted at in these macrophages. We first analyzed intracellular viral RNA (vRNA) levels using a flow-cytometric based RNA FISH assay (FISH-Flow). In this assay, 48 short (18-22 nt) DNA oligos were tiled across segment 2 of the influenza virus genomic RNA and labeled with a fluorescent fluorophore (ATTO 488). The collective accumulation of these probes binding to vRNA enable highly sensitive detection of vRNA positive cells. Influenza virus-infected macrophages were hybridized with FISH probes labeling viral RNA, and then the percent viral RNA positive cells were enumerated. Mock infected cells treated with FISH probes were used to generate a vRNA positive gate (Fig. 5.7C) and representative gates from influenza infected macrophages are shown in (Fig. 5.7D). We observed similar levels of vRNA-positive cells in iPSC- and blood-derived macrophages, and note that the levels of vRNA increased in an MOI-dependent manner (Fig 6D). We next assessed the expression levels of a viral protein, hemagglutinin (HA), at the cell surface. HA is the major virus surface glycoprotein that involved in cellular entry. Similar to what was observed for vRNA, we also observed significant numbers of HA-positive cells that also increased in an MOI-dependent manner (Fig. 5.7E). Similar experiments were performed in the highly permissive A549 cells, further validating these approaches (Fig. 5.8). Collectively, this data reveals that influenza virus infection is not likely restricted at the step of vRNA replication or protein translation, further suggesting that influenza virus production from macrophages is restricted during viral egress from the cell.

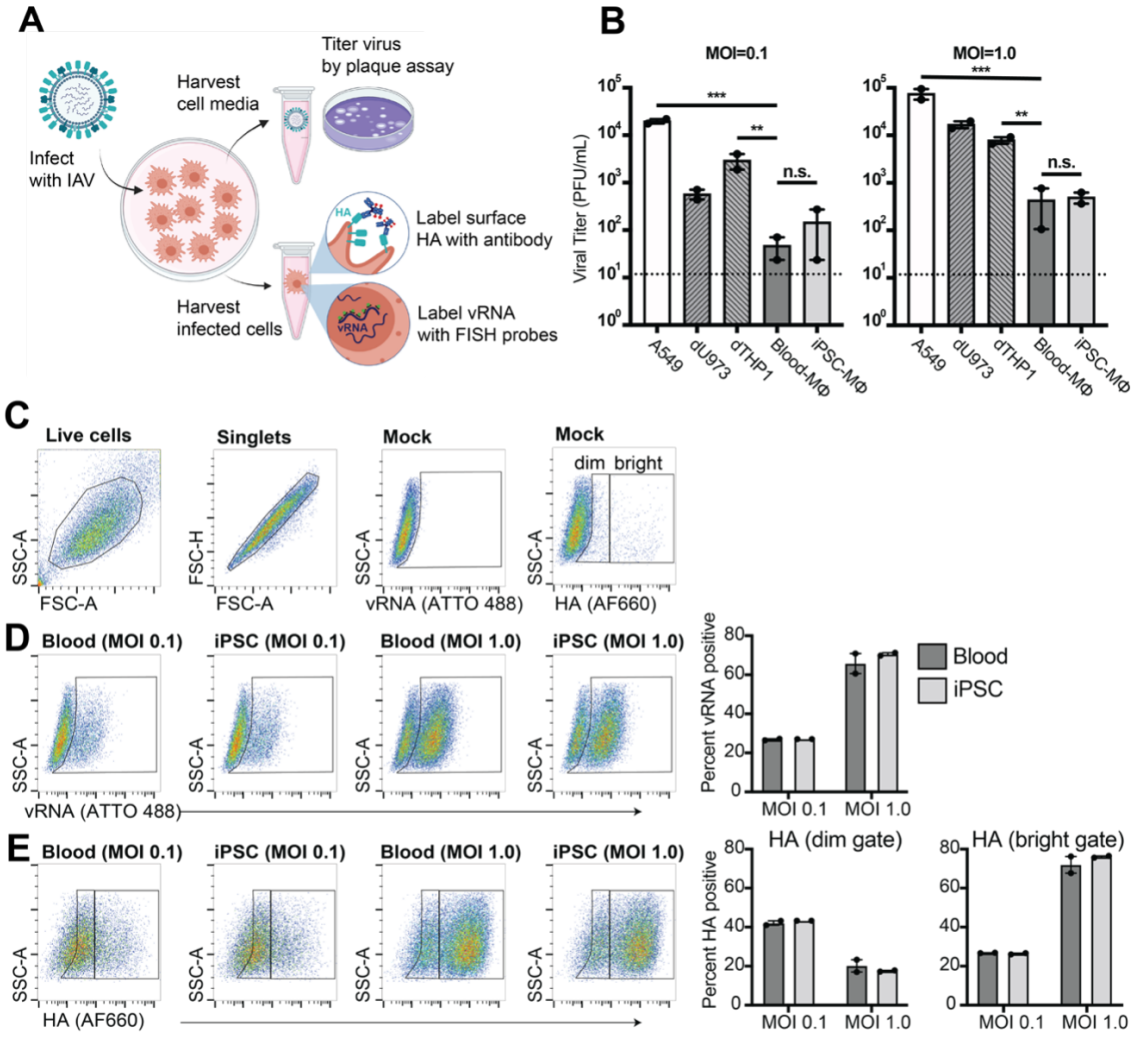


Figure 5.7: **Blood- and iPSC-derived macrophages exhibit similar restriction against influenza virus infection during virus budding.** **(A)** Schematics of experiment setup to quantify influenza virus genome replication, viral protein and infectious virion production. Blood/iPSC-derived macrophages, along with differentiated THP1, U937 and A549 cells were infected with influenza virus at indicated MOI of 0.1 and 1.0 for 24 hours. **(B)** The supernatant containing infectious virions were harvested and titered via plaque assay, while **(C-E)** intracellular viral RNA (vRNA) and surface hemagglutinin (HA) were analyzed by flow-cytometry . **(C)** Mock infected cells were used to draw vRNA- and HA-positive gates. For HA staining, both dim and bright gates were used to identify their respective HA-positive populations. **(D)** vRNA-positive and **(E)** surface HA protein-positive cells were enumerated by flow cytometry. All flow cytometric events were gated on forward area scatter (FSC-A) and side area scatter (SSC-A) properties, followed by singlet discrimination. Flow cytometric plots are representative of two independent experiments and data graphed display the mean +/- SEM from two biological replicates.

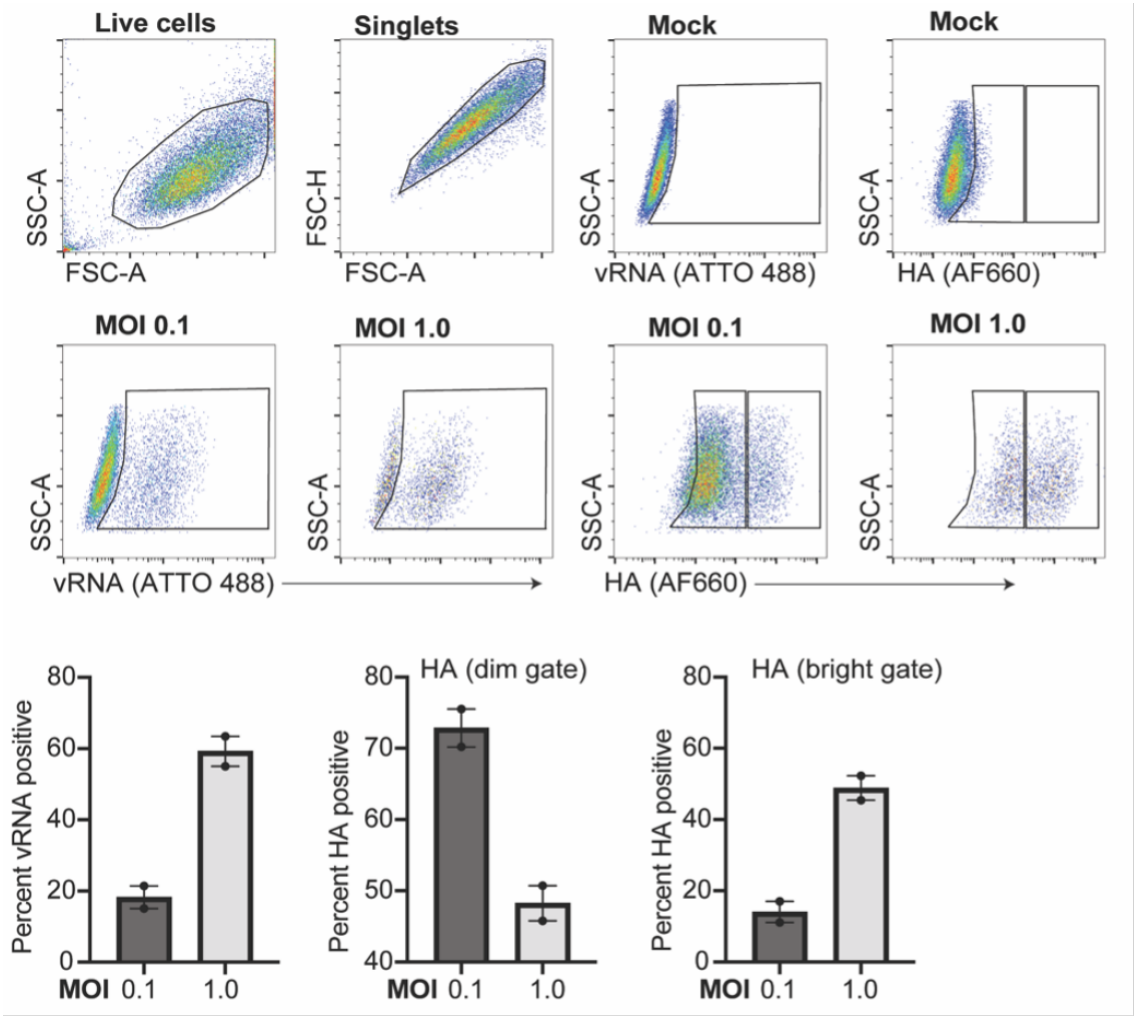


Figure 5.8: **Detection of vRNA and HA in influenza virus permissive A549 cells.** Mock infected cells were used to draw vRNA- and HA-positive gates. For HA staining, both dim and bright gates were used to identify their respective HA-positive populations. vRNA-positive and surface HA protein-positive cells were enumerated by flow cytometry. All flow cytometric events were gated on forward area scatter (FSC-A) and side area scatter (SSC-A) properties, followed by singlet discrimination.The flow cytometric plots from two independent experiments are shown and the corresponding bargraphs display the mean +/- SEM from two biological replicates.

5.5 Discussion

Monocytes and macrophages are essential cellular components of the innate immune system. They are oftentimes the first population of immune cells to encounter pathogens, and they play a critical role in microbial clearance through phagocytosis and activation of adaptive immunity. Despite their important role in immunity, these cells may also serve as a double-edged-sword; monocytes and macrophages are oftentimes exploited by viruses as vessels for viral replication and dissemination throughout the body. Thus, studying the biology of viruses in macrophages and how these essential innate immune cells respond to infection is of upmost importance.

The main source of monocytes and macrophages are primary cells derived from blood donors. However, despite their improved relevance over standard immortalized cell lines (THP-1, U937, etc.), there are numerous challenges that come along with working with these primary cell types. First, monocytes circulate in low numbers in the blood, cannot be propgated in vitro, and thus require large volumes of blood for routine experimentation. Second, the reliance on donors with different genetic backgrounds leads to heterogenetity in cellular response and function, which confounds experimental conclusions. To overcome these limitations, human iPSC-derived monocytes and macrophages have seen increased utilization. We routinely isolated monocytes every 4-5 days for up to a month from just a few wells of a 6-well dish. If greater cell yields were needed, one would just need to scale appropriately.

Using three clinically relevant human viruses, HIV-1, dengue virus, and influenza A virus, we show that iPSC monocyte-derived macrophages faithfully recapitulate relevant phenotypes observed in primary blood monocyte-derived macrophages. This is exemplified by several key findings. First, we demonstrate that HIV-1 and dengue virus replication are virtually identical in iPSC- and blood-derived macrophages. We observed similarities in viral growth kinetics and endpoint viral titers for dengue virus, and note that HIV-1 is similarly

capable of initiating cell-free and cell-to-cell transfer from persistently infected iPSC- and blood monocyte-derived macrophages. Second, we show that dengue virus replication is similarly restricted in iPSC and blood M1 macrophages, recapitulating a phenotype not seen in cell lines used in this field. Finally, we show that seasonal influenza A virus infection is similarly restricted in iPSC- and blood monocyte-derived macrophages at a late stage of the virus replication cycle, and that this blockade was not present in immortalized monocytic or epithelial cell lines. This later finding is significant because, despite the utility that PMA-induced THP-1 and U937 immortalized macrophages provide for influenza virus research, they do not faithfully recapitulate key viral restriction phenotypes that are observed in primary human cells. When taken together, iPSC-derived monocytes and macrophages should be considered a vital cell culture model system for studying virus biology and host response to infection.

In summation, we have shown that iPSC-derived monocytes and macrophages are a valuable model system for probing questions related to virus biology and host response to infection. In addition, we have provided one of the most comprehensive comparisons of global gene expression and genome structure changes during iPSC and blood cell differentiation into monocytes and M0, M1, and M2 monocyte-derived macrophage subsets. We conclude that iPSC-derived cells display near identical properties to primary blood-derived cells and thus would serve as a valuable alternative to primary cells obtained from human blood donors.

5.6 Method

5.6.1 Cell lines and culture conditions

Monocytic cell lines THP-1 (ATCC TIB-202) and U937 (ATCC CRL-1593.2) cells were cultured in RPMI-1640 Medium (ATCC #30-2001) with 10% FBS and 1% Pen Strep (complete medium). TZM-bl (NIH ARP # 8129), MDCK (ATCC CCL-34) and BHK (ATCC PTA-3544) cells were cultured in Dulbecco's modified Eagle medium (Invitrogen) with 10%

FBS, 2 mM L-glutamine, and 1% Pen Strep (complete medium). A549 cells were cultured with F-12K medium supplemented with 10% FBS, 2 mM L-glutamine, and 1% Pen Strep. The iPSC cell line, iC7-2, was obtained from University of Colorado Anschutz Gates Center for Regenerative Medicine, and the iCTR was obtained from Cedars-Sinai Medical Center iPSC Core. Specifically, iC7-2 line was generated from deidentified specimens from Lonza's publicly available biorepository (Lonza C7 fibroblasts). The reprogramming was carried out by the Gates Center for Regenerative Medicine staff via transfection of modified mRNAs and miRNAs using a virus-free and non-integrating method[220]. The iCTR cell line was generated from healthy donor PBMCs (CS0594iCTR), and the reprogramming was performed via nucleofection of episomal plasmids by the core staff as previously described[221]. The iPSC lines were characterized through karyotype analysis, mycoplasma testing and pluripotency testing. Both iPSC lines are maintained in mTeSR Plus medium (STEMCELL #100-0276) using hESC-qualified Matrigel (Corning CLS354277) coated plates. All cells were maintained at 37C and 5% CO2.

5.6.2 Differentiation of iPSC-derived monocytes using STEM Diff

For each iPSC cell line, approximately 120 iPSC aggregates with 50-200 μm were seeded in 4 wells of hESC-qualified Matrigel (Corning CLS354277) coated 6-well tissue culture plates. The next day, the plates were examined under a light microscope to ensure around 60-80 iPSC aggregates were seeded within each well of the 6-well plate. The differentiation of iPSC into monocytes followed the STEMdiff Monocyte Kit protocol specified by the manufacturer (STEMCELL # 05320). By day 10 of the protocol, monocytes were observed to lift off the basal hematopoietic progenitor cells and remain in the media suspension. Subsequently, we followed a fixed schedule to harvest monocytes from the culture supernatant: starting with 2 mL of monocyte differentiation medium (StemSpan SFEM II [STEMCELL #09605] + 1X STEMdiff monocyte differentiation supplement [STEMCELL #05324]) in each well, we incubate the culture for 2 days; we then top up with 2 mL addi-

tional monocyte differentiation medium and incubate for another 2-3 days; finally, we harvest the entire supernatant containing differentiated monocytes after 4-5 days, and replace with 2 mL fresh monocyte differentiation medium. This harvest schedule was repeated 7 times over the course of 6 weeks.

The harvested monocytes in the supernatant were collected via centrifugation at 300xg for 5 minutes. And a media transition technique was adapted to introduce the iPSC-derived monocytes to the ImmunoCult-SF Macrophage Medium (STEMCELL #10961). Specifically, medium ratio of 1:3 for monocyte differentiation medium:ImmunoCult-SF macrophage medium was used to plate the harvested monocyte onto fresh 6-well tissue culture plates on the day of monocyte harvest. In the following days, a daily media change was performed on the adherent monocytes following the monocyte differentiation medium:ImmunoCult-SF macrophage medium ratio of 2:2 (1 day after harvest) and 3:1 (2 days after harvest). We started iPSC-derived monocyte to macrophage differentiation 3 days after monocyte harvest by replacing medium with 100% ImmunoCult-SF macrophage medium containing 10 ng/mL M-CSF.

5.6.3 Isolation of blood-derived monocytes

To obtain blood-derived monocytes, 50 mL of human blood was collected on the day of purification. Within 1 hour of blood collection, 12 mL aliquots of blood were diluted 1:1 with PBS + 2% FBS before being loaded onto 15 mL of Lymphoprep (STEMCELL #07801). The layered blood-Lymphoprep mixture was centrifuged at 800xg for 20 minutes at room temperature with brake off. After the centrifugation, the plasma layer was gently removed and discarded. The mononuclear cell layer was collected with washed with PBS + 2% FBS for at least three time with low-speed centrifugation at 180xg for 10 minutes at room temperature to remove the contaminating platelets. Subsequently, the monocytes were purified from mononuclear cells via Percoll (Sigma-Aldrich P4937) gradient centrifugation following published protocol (Mosher 2021). Briefly, mononuclear cells from preview step

were resuspended in 3 mL of RMPI+1%FBS and slowly layered on top of density gradient. The density gradient was prepared within a 15-mL conical tube: First, a 100% Percoll fraction was prepared by mixing 4.5 mL Percoll with 0.5 mL 1x dPBS; then the 52.5% Percoll fraction was prepared by mixing 2.675 mL 100% Percoll fraction with 2.325 mL RPMI+1%FBS, and the 45% Percoll fraction was prepared by mixing 2.25 mL 100% Percoll fraction with 2.75 mL RPMI+1%FBS; finally, the density gradient was prepared by slowly layering 4.5 mL of the 52.5% Percoll fraction followed by 4 mL of 45% Percoll fraction. The density gradient containing mononuclear cells was then centrifuged at 610xg for 30 minutes at room temperature with the brake off. The middle layer, containing monocytes, were collected and washed with PBS via centrifugations at 250 xg for 10 minutes at room temperature three times. The final monocytes were plated onto 6-well plate at $2x10^6$ cells per well.

5.6.4 Monocyte to macrophage differentiation

To differentiate monocytic cell lines into macrophages, THP-1 and U937 cells were plated at $2x10^6$ cells per well of a 6-well dish in RPMI-complete medium supplemented with 10 ng/mL PMA (phorbol 12-myristate 13-acetate; Sigma-Aldrich P8139) and cultured for 3 days. To differentiate both iPSC-derived and blood-derived monocytes into naïve macrophages, monocytes were plated at $1x10^6$ cells per well of a 6-well in ImmunoCult-SF macrophage medium supplemented with 10 ng/mL M-CSF (Fisher Scientific 216-MC-010) and cultured for 4 days. To polarize the naïve macrophages into M1 subtype, the M0 macrophages were incubated in ImmunoCult-SF macrophage medium supplemented with 50 ng/mL Interferon-gamma (R&D 285-IF-100/CF) and 10 ng/mL lipopolysaccharide (Sigma-Aldrich L2018) for 2 days. To polarize the naïve macrophages into M2 subtype, the M0 macrophages were incubated in ImmunoCult-SF macrophage medium supplemented with Interlukin-4 (Sigma-Aldrich H7291) for 2 days.

5.6.5 Phenotypic profiling of surface markers by flow cytometry

To carry out phenotypic profiling, at least 1x10^5 cells was lifted off the tissue culture plate via treatment of Accutase (STEMCELL #07922). The cells were washed once with PBS, and incubated with 25 μg/mL Fc receptor block (BD #564220) diluted in 50 μL FACS buffer (PBS + 1mM EDTA + 2% FBS) for 10 minutes at room temperature. Without removing Fc blocking solution, additional 1 uL of corresponding stock fluorescently labelled cell surface antibodies were spiked into the solution (CD14-PE/Cy7 [BioLegend 301813], CD11b-PE [BioLegend 301305], CD80-PE [BioLegend 305207], CD206-FITC [BioLegend 321113],). The cells are then incubated in dark at 4C for 30 minutes to allow antibody binding. After the incubation, unbound antibody was washed off twice using FACS buffer. The flow cytometry analysis was carried out using BD Accuri C6 Plus flow cytometer, and the analysis was done using FlowJo analysis software v10.8.0.

5.6.6 Transcriptomic profiling using RNA-seq and ATAC-seq

ATAC-seq and RNA-seq Library preparation. iPSCs, along with blood- and iPSC-derived monocytes/macrophages are cultured in 6-well plates at 1x10^6 cells/well, each with biological duplicates. The cells were lifted via treatment of Accutase (STEMCELL #07922) for 20 minutes at 37 C, and 5x10^4 cells were split for ATAC-seq library preparation following published Omni-ATAC protocol [11]. The ATAC-seq library was further size-selected using BluePippin machine with 2% agarose gel cassette (Sage Science RBF2010) to enrich for 100-1000 bp fragments. For RNA-seq, total RNA was harvested using Zymo Quick-RNA Microprep kit (Zymo R1050), and mRNA was enriched and prepared into sequencing library using KAPA mRNA HyperPrep kits (Roche #8098123702). Both ATAC-seq and RNA-seq were sequenced on NovaSeq 6000 System using 2x151 cycles for a minimum of 20 million pair-end reads/library.

RNA-seq data analysis. Raw sequencing reads were first trimmed for low-quality reads

and adapter sequences using BBDuk (BBMap v38.05). The trimmed reads are mapped to human hg38 genome using HISAT2 v2.1.0. The number of reads mapped to human RefSeq exons using FeatureCount (Subread v1.6.2). And the reads were normalized by size factors and analyzed for differential expression using DESeq2 R package v1.34.0. To visualize relative expression of genes, the fragments per kilobase of transcript per million mapped reads (FPKM) of size-factor normalized read counts were extracted to plot the heatmaps, and the expression level was subsequently scaled to row z-score. The gene ontology enrichment analysis was done using R package clusterProfiler v3.0.4.

ATAC-seq data analysis. Raw sequencing reads were first trimmed for low-quality reads and adapter sequences using BBDuk (BBMap v38.05). The trimmed reads are mapped to human hg38 genome using Bowtie2 v2.2.9. The resulting mapped reads were used to identify enriched open-chromatin peaks using the software package Genrich v0.6.1 in ATAC-seq mode, which automatically removes mitochondrial reads, PCR duplicates, and correct for read shifting caused by Tn5 transposition (Berg et al 1983). The identified narrow peaks were then merged across biological duplicates and experimental conditions. The number of reads mapped to each peak were counted using FeatureCount (Subread v1.6.2). The read counts were then corrected for library size and Trimmed Mean of M-values (TMM) using edgeR v3.36.0. We then identified open chromatin regions that underwent significant accessibility changes during differentiation and/or polarization and clustered their accessibility profiles using hierarchical clustering. The motifs that are significantly enriched within each cluster were identified using Analysis of Motif Enrichment (AME, MEME suite 5.4.1) against the reference human motif database HOCOMOCO v11.

5.6.7 Virus production, macrophage infections, and titering

Dengue virus. Propagation of Dengue virus. C6/36 mosquito cells were seeded at $1x10^5$ cells/cm2 in a 150 cm2 flask to reach 80% confluency the next day. To make infection medium, Dengue virus stock (DENV2 Thailand/16681/84) was diluted in PBS with calcium

and magnesium to reach MOI of 0.001 in a total volume of 3 mL. The cells were incubated with virus inoculum at 28°C for 1 hour. After the incubation, the inoculum was removed, and cells were washed once with PBS. The cells were then incubated in EMEM+HEPES+2%FBS for 7 days when additional fresh medium was added. On day 13, the supernatant from infection was collected, and cell debris were spun down at 1,000xg for 5 minutes. The titer of the dengue virus stock was determined using plaque assay on BHK cells as described below.

Infection of macrophage. Macrophages were seeded at $1x10^6$ cells/well in 6-well plates on the day before infection. Dengue virus was diluted in ImmunoCult-SF macrophage medium at the indicated MOI and incubated with macrophages for 1 hour. After the incubation, the incoulumn was removed, and the cells were washed with PBS three times. Finally, 2 mL ImmunoCult-SF macrophage medium was replaced on the cells. The supernatant was collected at the indicated time after infection.

Infectious virus determination via the plaque assay. The titer of the dengue virus stock as well as supernatant from macrophage infections were determined with plaque assays on BHK cells. BHK cells were seeded into 6 well-plates at $1x10^6$ cells/well to reach near confluency the next day. The virus stock or the infection supernatant were 10-fold serial diluted in serum free DMEM, and 400 μL of the dilution was overlayed on BHK cells for 1 hour at 37°C. After 1 hour, the inoculum was replaced with EMEM containing 1.2% Avicel and 10% FBS and incubated at 37°C for 5 days. After the incubation, the overlay medium was washed off with PBS, the cells were fixed and visualized with 20% methanol containing 0.2% crystal violet. The number of plaques were manually counted and reported as PFU/mL.

HIV-1. Generation of virus stocks. Blood was obtained from a healthy donor and PBMCs were isolated following density gradient centrifugation (Lymphoprep; STEMCELL Technologies cat# 07811). The isolated PBMCs were then cultured in RPMI complete media supplemented with 5 μg/mL PHA (phytohemagglutinin; Sigma-Aldrich # 11249738001) and 20 U/mL IL-2 (interleukin-2; PeproTech #200-02) for 3 days. After 3 days of PHA

stimulation, the media was removed and the cells were expanded in RPMI complete plus IL-2 (T cell maintenance medium). To prepare HIV-1 stocks, $1x10^7$ PBMCs were pelleted at 300 x g for 8 min, the media was removed, and the cell pellet was then resuspended in 1 mL of HIV-1 SF162 cell-free virus (NIH ARP #276). After a 1h incubation with virus, the culture volume was increased to 10 mL (T cell maintenance medium) and the virus culture was maintained for 14 days with the following considerations: on day 3, half of the media was replaced with fresh media; on day 7, the culture volume was doubled by adding $1x10^7$ fresh PBMCs; on day 9, half of the media was replaced with fresh media; and on day 14, the cell supernatant was harvested and titered by TCID50 assay on TZM-bl reporter cells.

Macrophage infections. iPSC (n = 2 cell lines) and blood-derived M0 macrophages (n = 1 donor in 3 biological replicates) were plated at $1x10^5$ cells per well of a 48 well dish. The next day, the cells were infected with HIV-1 SF162 virus ($3.2x10^5$ TCID50/mL) in the presence of 5 μg/mL polybrene. The cells were spinoculated at 1200 x g for 75 mins at 30C. Following spinoculation, the cells were washed 3X with PBS and resuspended in 500 μl of macrophage media (ImmunoCult-SF Macrophage Medium; STEMCELL Technologies cat# 10961). After 7 days of culture, the media was removed and stored at -80C for virus titering.

Infectious virus determination. Macrophage cell supernatants were split into 3 equal volumes (technical replicate samples) and were inoculated onto TZM-bl indicator cells seeded the day prior at $1x10^4$ cells/well of a 96-well dish. TZM-bl indicator cells stably express the HIV-1 receptors CD4 and CCR5, and contain an integrated reporter gene for firefly luciferase whose expression is activated by HIV-1 gene products. So, luciferase activity from these cells indicates productive viral infection. The cells were spinoculated in the presence of 5 μg/mL polybrene as described above, and then allowed to incubate for 48h at 37C, 5% CO2. After the 48h incubation, the media was removed, the cells were washed, and then virus infectivity was measured by firefly luciferase assays following the manufacturers protocol (Promega). Luminescence was determined using the Synergy LX plate reader (Biotek). Macrophage-TZM-bl coculture. Following 7 days of culture, infected macrophages (described

above) were washed, detached from the culture dish using Accutase cell detachment solution (STEMCELL Technologies cat# 07922), split into two equal volumes (technical replicate samples) and plated onto TZM-bl indicator cells seeded the day prior at $1x10^4$ cells/well of a 96-well dish. The macrophages and TZM-bl cells were co-cultured for 3 days. The transmission of infectious virus from macrophages to naïve TZM-bl cells was measured by luciferase assay as described above.

Influenza virus. Generation of viral stocks. Influenza H3N2 influenza A Udorn virus stocks (Chen et al., 2007) were grown in 10-day-old fertilized chicken eggs, and virus stock was further propagated using human A549 cells. Specifically, A549 cells were plated on 15 cm dish at $1.5x10^7$ cells to reach 100% confluency. The next day, $1.5x10^5$ PFU viral stock was diluted in 5 mL serum free DMEM medium (MOI = 0.01) for 1 hour infection at 37C. Following the incubation, the infection medium was replaced with 20 mL DMEM + 0.3% Bovine Serum Albumin + 1.0cμg/ml N-acetylated trypsin for 72 hours. The supernatant containing infectious viruses were collected, centrifuged at 500xg for 10 minutes to get rid of cell debris. The titer of the influenza virus stock was determined via plaque assay on MDCK cells.

Infection of macrophage, A549. Macrophages were seeded at $1x10^6$ cells/well in 6-well plates on the day before infection. Influenza virus was diluted in ImmunoCult-SF macrophage medium at the indicated MOI and incubated with macrophages for 1 hour. After the incubation, the incoulumn was replaced with 2 mL ImmunoCult-SF macrophage medium containing 0.2 cμg/mL N-acetylated trypsin for 48 hours. In the case of immortalized macrophage infection (dTHP1, dU937), serum-free RPMI was used to dilute influenza virus stock, and serum-free RPMI 0.2 + 0.2 cμg/mL N-acetylated trypsin + 0.3% BSA was used to incubate the infected cells after 1 hour virus adherence. To infect A549 cells, influenza virus was diluted in serum-free F-12K medium and incubated with the cells for 1 hours before being replaced with 2 mL serum-free F-12K medium + 0.2 cμg/mL N-acetylated trypsin.

Infectious virus determination via the plaque assay. The titer of the influenza virus stock as well as supernatant from macrophage/A549 infections were determined with plaque assays on MDCK cells. MDCK cells were seeded into 6 well-plates at 1x10^6 cells/well to reach near confluency the next day. The virus stock or the infection supernatant were 10-fold serial diluted in serum free DMEM, and 850 μL of the dilution was overlayed on MDCK cells for 1 hour at 37°C. After 1 hour, the inoculum was replaced with MEM containing 1.2% Avicel and incubated at 37°C for 48 hours. After the incubation, the overlay medium was washed off with PBS, the cells were fixed and visualized with 20% methanol containing 0.2% crystal violet. The number of plaques were manually counted and reported as PFU/mL.

HA-staining and FISH-flow for influenza virus infections. To carry out quantification of influenza viral protein and viral RNA via flow cytometry, at least 1x10^6 mock/infected cells was lifted off the tissue culture plate via treatment of Accutase (STEMCELL #07922). The cells were washed once with PBS and incubated with 25 μg/mL Fc receptor block (BD #564220) diluted in 50 μL FACS buffer (PBS + 1mM EDTA + 2% FBS) for 10 minutes at room temperature. Without removing Fc blocking solution, additional 1 μL of anti-H3N2 HA mouse monoclonal antibodies were spiked into the solution (1:50 dilution, Sino Biological #11056). The cells are then incubated in dark at 4C for 30 minutes to allow antibody binding. After the incubation, unbound antibody was washed off twice using FACS buffer. We then applied the secondary Alexa-647 goat anti-mouse antibodies diluted 1:100 in FACS buffer to the cells and incubated in dark at 4C for 30 minutes. The antibody labelled cells were washed twice more using FACS buffer, followed by fixation and permeabilization using BD Cytofix/Cytoperm kit (BD BDB554714). To visualize the viral RNA, cells were washed once with Wash Buffer A (Biosearch, #SMF-WA1-60) and resuspended in 50 μL of hybridization buffer containing FISH probes (1:100 dilution, Biosearch Technologies, #SMF-HB1-10). The cells were incubated in the dark at 37C for 6 hours. After incubation, cells were washed twice using Wash Buffer A, incubated at 37 C for 30 minutes, washed once more with Wash Buffer B (Biosearch Technologies, #SMF-WB1-20). Finally, the cells were resuspended in FACS

buffer for analysis. The flow cytometry analysis was carried out using BD Accuri C6 Plus flow cytometer, and the analysis was done using FlowJo analysis software v10.8.0

5.6.8 Human subjects for blood donation

Blood samples used in this research were obtained from anonymous individuals consented under human study #20-0068, approved by the University of Colorado Institutional Review Board. Potential participants were verbally screened for their ability to meet inclusion and exclusion criteria. There were minimal inclusion criteria: subjects must be adults with the ability to consent and ability to provide a sample. Potential subjects were excluded if they reported a body weight of less than 110lbs or current pregnancy. Fourteen subjects enrolled and consented to blood draws for this study. This does not represent 14 unique individuals as subjects were able to enroll more than once. No data were collected from subjects to maintain anonymity and confidentiality. The IRB approved a waiver of written consent as this was deemed a minimal risk study and the only record linking the subject to the research would be their name on a consent form, leading to a possible harm if confidentiality was breached. No financial or non-financial compensation was given to subjects for participation.

Chapter 6

Conclusions and Future Directions

6.1 Conclusions

Like many of my peers, my graduate study was heavily shaped by the COVID-19 pandemic. In this dissertation, I described several projects that address both infectious disease surveillance and modeling in response to the pandemic. By analyzing the on-campus SARS-CoV-2 screening data, I realized that asymptomatic individuals could harbor just as much viruses as the symptomatic patients, and that the asymmetrical distribution of circulating viruses within the community suggests the majority of transmission events are likely resulted from a few virus super-carriers (Chapter 2).Such transmissions are likely happening unnoticed within the asymptomatic population, making contact tracing difficult. This further highlights the importance of using frequent and accessible infection screening test to identify these contagious yet asymptomatic individuals. This first drove me to develop a rapid saliva-based SARS-CoV-2 test (Chapter 3). The use of noninvasive saliva samples in this test lowered the behavior barrier that deters seemingly healthy individuals from getting regular tests. And the simplicity of the test itself helped bypass supply chain issues around the standard RT-qPCR-based diagnostics used during the pandemic, allowing for frequent community-wide SARS-CoV-2 screenings. Subsequently, to extend the idea of asymptomatic population screening beyond the current pandemic, I examined the host immune response for signatures that signal ongoing infection (Chapter 4). I found a group of mRNA transcripts derived from the human immune response to be highly dependable in distinguishing

in vitro infection conditions (mock/infected), and further validated their presence in infected human saliva samples. Through additional validation using clinical saliva samples derived from individuals infected with diverse pathogens, I showed this host response-based infection screening strategy has the potential to identify diverse infections before the onset of symptoms, further addressing the overarching goal of infectious disease surveillance.

Finally, to provide a more tractable model for future host immune response studies, I hence validated an iPSC-derived macrophage model system in the context of viral-host interactions (Chapter 5). The validation was carried out between blood-derived and iPSC-derived monocytes/macrophages at the phenotypical level that looked at surface protein expression, at the transcriptional level that compared transcriptomes and the chromatin accessibility, and at the functional level that contrasted the susceptibility to multiple viral infections. Through these validations, I found that the iPSC-derived macrophages can serve as a relevant alternative model to the blood-derived counterpart, with the advantages of being readily obtainable in large numbers and amendable to genome editing. This opens avenues of future research that characterize of macrophage-tropic viral infections, profile the transcription regulation during macrophage maturation, or compare population-wide immune response differences.

6.2 Future Directions

6.2.1 Macrophage-specific restriction factors against viral infection

Macrophage as an immune cell type is known to express a diverse set of anti-viral restriction factors that directly binds to the viral genome or proteins and inhibit the viral replications. In fact, during the functional validation of iPSC-derived macrophages, we already noticed a significant restriction on dengue virus replication by M1 macrophages (see Figure 5.6), as well as on seasonal influenza virus replication by naive macrophages (see Figure 5.7). While numerous interferon-induced host immune proteins have already

been identified to restrict dengue virus (TRIM69, STING, RTP4, IFITM3) and influenza virus (TRIM25, ISG15, RSAD2) replication [222–227], the possibility of macrophage-specific restriction still exists, as macrophage is under drastically different transcriptional regulation that leads to expression of a more comprehensive collection of immune genes, compared to other somatic cell types used by previous studies [228].

To identify the macrophage-specific restrictions against viruses, a gain-of-function genome-wide cDNA screen technique can be used (Figure 6.1). This is achieved by purifying total RNA from primary monocyte-derived M1 macrophages. As the M1 macrophages are polarized towards the inflammatory subtype using LPS and IFN-*gamma*, it is reasonable to believe that M1 macrophages would harbor the most comprehensive immune response RNA transcripts. Using poly-dT beads, the mRNA was hence captured from the total RNA, representing the M1 macrophage transcriptome. The purified mRNA is then reverse transcribed into cDNA containing gateway cloning adapter sequences. This allowed flexible shuttling of the entire macrophage cDNA library across different cloning vectors. In this case, the gateway vector containing macrophage cDNA can then be shuttled into a lentiviral transduction system to generate macrophage cDNA-containing lentiviruses using 293T cells. This lentiviral library can then be applied on to susceptible cells that would usually be lysed during the infection by the virus-of-interest. In the presence of protective genes derived from M1 macrophages, we would expect the survival of certain transduced cells during the viral infection. The survivor cells can then be compared to the transduced but uninfected cells to reveal which macrophage cDNA provided protection against the viral infection. Using this approach, I hope to identify restriction factor(s) that provide(s) protection in susceptible cells (A549, Huh7, Vero cells) in the context of influenza virus or dengue virus infection. The use of such genome-wide gain-of-function cDNA screen has the potential for the fast and reliable identification of multiple antiviral restriction factors that will significantly aid the future understanding of host-pathogen interactions.

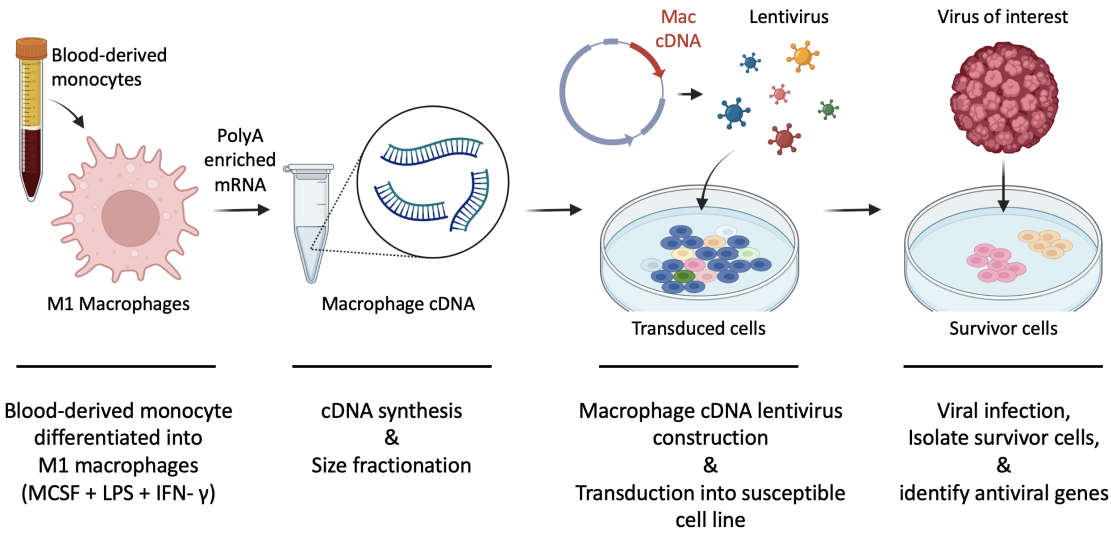


Figure 6.1: **Overview of the genome-wide macrophage cDNA screening process.** To carry out gain-of-function viral restriction factor screening, mRNA from primary monocyte-derived M1 macrophages was extracted via poly-dT capture, and reverse transcribed into cDNA containing gateway cloning adapter sequences. The gateway vector containing macrophage cDNA is then shuttled into lentiviral transduction system. This lentiviral system is then transfected into human 293T cells to generate lentiviruses containing macrophage cDNA. This lentiviral macrophage cDNA library is then applied on to susceptible cells that would usually undergo cytopathy during the infection by the virus-of-interest, but would survive due to the presence of the additional protective macrophage cDNA. The survivor cells were then compared to the initially transduced cells to reveal which macrophage cDNA provided protective affect against the viral infection.

6.2.2 Immune transcription regulation during iPSC-Macrophage differentiation

Despite the essential functions of innate immune response in restricting infections, it is previous shown that, in stark contrast to differentiated cells, pluripotent stem cells lack the capacity to respond to pathogen associated molecular patterns and are unable to initiate the subsequent interferon response [229]. This observation is linked to the combined effects of low expressions of pattern recognition receptors, as well as the inhibition of interferon stimulated transcription factors [229]. However, the underlying reason of such interferon-inert phenotype, as well as the transcriptional regulation events that initiated the interferon response downstream of pluripotent stem cell differentiation remain unknown. One hypothesis is that the transcription factor that defines pluripotency, Kruppel Like Factor 4 (KLF4), likely represses immune activation through direct binding to interferon-stimulated response elements in the genome [229]. During iPSC differentiation, the KLF4 is not only important for mesoderm lineage commitment, but also for monocyte differentiation as well as macrophage polarization [230–233]. This makes the iPSC-macrophage lineage extremely relevant to study the immune transcription regulation during the stem cell differentiation. Due to the heterogeneous nature of stem cell differentiation, future characterization should focus on the transcription profile at the single-cell level, specifically via combined single-cell RNA-seq and single-cell ATAC-seq. This combined profiling would help answer questions such as at what stage of differentiation the cells become responsive to innate immune stimulation, and what are the essential transcription factors that define such stage of differentiation. Our preliminary work has indicated key transcription regulation events that happened during the monocyte-to-macrophage subtypes differentiation/polarization (see Figure 5.4), the follow-up studies should focus on the differentiation stages prior to monocytes, as the myeloid cells are already mature in immune response.

Additional complexity to this matter comes from the observation that individuals with

trisomy chromosome 21, clinically diagnosed as Down syndrome, also show distinct composition of blood cell types comparing to healthy individuals [234]. This lead to the hypothesis that the presence of additional copy of chromosome 21, which harbors a large set of immune-regulatory transcription factors and immune genes, could affect the differentiation of certain blood cell types during development [235]. This is further supported by the finding that runt-related transcription factor 1 (RUNX1) not only displays abnormal activation in trisomy 21 cell lines, but it also plays an important role in myeloid cell line differentiation (Figure 5.4) [236, 237]. Using the iPSC-to-macrophage model system described herein, future studies should investigate the differentiation capacity of trisomy 21 iPSCs into myeloid cell types. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) targeting the key transcription factors that are affected by trisomy 21, such as RUNX1, could be used to profile such transcription factor activity over the course of iPSC-to-macrophage differentiation. And such ChIP-seq comparison between trisomy and disomy 21 cell lines could hence help reveal the mechanism underlying the trisomy 21-specific hematopoiesis deficits.

6.2.3 Population-wide immune response differences

Finally, one fascinating aspect of the human immune response is the population-level diversity. As different groups of human ancestors venture out to different corners of the world, they encounter distinct pathogens that they eventually adapted to. On the other hand, the loss of selective pressure from pathogens renders certain ethnic group capable of eliciting stronger immune response than the other [184]. Previous studies have found such difference in immune response can be attributed to population wide genetic polymorphisms, and such polymorphism exists largely outside of the protein coding regions of our genome. Via expression quantitative trait loci (eQTL) analysis, previous studies have narrowed such genetic diversity to regions of our genome that are linked to differential expressions of immune related genes. This leads to the hypothesis that such genetic diversity exists within the gene regulatory regions, such as enhancers and promoters[184]. The existence of single nu-

cleotide polymorphism within these regulatory regions could hence affect the transcription binding activity, which subsequently modulate the differences in immune gene expression upon activation. To profile this intricate relationship, nascent RNA sequencing, such as precision run-on sequencing (PRO-seq), can be used to monitor the genome-wide enhancer activation upon immune activation via interferon treatment or viral infections [238]. Future studies should utilize cell lines that represent the population diversity across different ethnic groups, such as the lymphoblastoid cells available from 1000 Genome project [239] or iPSC derived from diverse individuals. Such studies can then use PRO-seq to identify polymorphic immune regulatory regions that exhibit differential activation across individuals. These findings can be cross referenced with the existing eQTL data to link polymorphic regulatory regions to immune genes that are deferentially expressed, and with the existing CHIP-seq dataset to reveal likely affected transcription factors. Together, a comprehensive understanding on how genetic diversity at immune regulatory regions attributes to immune response differences can advance our knowledge of evolution outside of protein coding regions, and can potentially link existing immune diseases with genetic variants.

Bibliography

[1] CDC. **Current Outbreak List**. Feb. 2022. URL: https://www.cdc.gov/outbreaks/index.html (visited on 02/22/2022).

[2] Tim R. Mercer and Marc Salit. "Testing at scale during the COVID-19 pandemic". In: **Nature Reviews Genetics** 22.7 (2021), pp. 415–426. ISSN: 1471-0056. DOI: 10.1038/s41576-021-00360-w.

[3] Markus Hoffmann et al. "Chloroquine does not inhibit infection of human lung cells with SARS-CoV-2". In: **Nature** 585.7826 (2020), pp. 588–590. ISSN: 0028-0836. DOI: 10.1038/s41586-020-2575-3.

[4] Jia Liu et al. "Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting SARS-CoV-2 infection in vitro". In: **Cell Discovery** 6.1 (2020), p. 16. ISSN: 2056-5968. DOI: 10.1038/s41421-020-0156-0.

[5] Bruce Alberts et al. **Molecular Biology of the Cell**. Garland Science, 2007. DOI: 10.1201/9780203833445.

[6] Leonidas C Platanias. "Mechanisms of type-I- and type-II-interferon-mediated signalling". English. In: **Nature Reviews Immunology** 5.5 (2005), pp. 375–386. ISSN: 1474-1733. DOI: 10.1038/nri1604.

[7] Anette H.H. van Boxel-Dezaire, M.R. Sandhya Rani, and George R. Stark. "Complex Modulation of Cell Type-Specific Signaling in Response to Type I Interferons". In: **Immunity** 25.3 (2006), pp. 361–372. ISSN: 1074-7613. DOI: 10.1016/j.immuni.2006.08.014.

[8] David R Bentley et al. “Accurate whole human genome sequencing using reversible terminator chemistry”. English. In: **Nature** 456.7218 (Nov. 2008), pp. 53–59. ISSN: 0028-0836. DOI: 10.1038/nature07517.

[9] Zhong Wang, Mark Gerstein, and Michael Snyder. “RNA-Seq: a revolutionary tool for transcriptomics”. In: **Nature Reviews Genetics** 10.1 (2009), pp. 57–63. ISSN: 1471-0056. DOI: 10.1038/nrg2484.

[10] Michael I Love, Wolfgang Huber, and Simon Anders. “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2”. In: **Genome Biology** 15.12 (2014), p. 550. ISSN: 1465-6906. DOI: 10.1186/s13059-014-0550-8.

[11] M Ryan Corces et al. “An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues”. In: **Nature Methods** 14.10 (2017), pp. 959–962. ISSN: 1548-7091. DOI: 10.1038/nmeth.4396.

[12] Zhijian Li et al. “Identification of transcription factor binding sites using ATAC-seq”. In: **Genome Biology** 20.1 (2019), p. 45. ISSN: 1474-7596. DOI: 10.1186/s13059-019-1642-2.

[13] Fangjie Zhu et al. “The interaction landscape between transcription factors and the nucleosome”. In: **Nature** 562.7725 (2018), pp. 76–81. ISSN: 0028-0836. DOI: 10.1038/s41586-018-0549-5.

[14] Daisuke Hirayama, Tomoya Iida, and Hiroshi Nakase. “The Phagocytic Function of Macrophage-Enforcing Innate Immunity and Tissue Homeostasis”. In: **International Journal of Molecular Sciences** 19.1 (2017), p. 92. DOI: 10.3390/ijms19010092.

[15] Guillermo Arango Duque and Albert Descoteaux. “Macrophage Cytokines: Involvement in Immunity and Infectious Diseases”. In: **Frontiers in Immunology** 5 (2014), p. 491. ISSN: 1664-3224. DOI: 10.3389/fimmu.2014.00491.

[16] Peter J. Murray et al. “Macrophage Activation and Polarization: Nomenclature and Experimental Guidelines”. In: **Immunity** 41.1 (2014), pp. 14–20. ISSN: 1074-7613. DOI: 10.1016/j.immuni.2014.06.008.

[17] Marco Orecchioni et al. “Macrophage Polarization: Different Gene Signatures in M1(LPS+) vs. Classically and M2(LPS–) vs. Alternatively Activated Macrophages”. In: **Frontiers in Immunology** 10 (2019), p. 1084. DOI: 10.3389/fimmu.2019.01084.

[18] Vincenzo Torraca et al. “Macrophage-pathogen interactions in infectious diseases: new therapeutic insights from the zebrafish host model”. In: **Disease Models & Mechanisms** 7.7 (2014), pp. 785–797. ISSN: 1754-8403. DOI: 10.1242/dmm.015594.

[19] Peter J. Murray and Thomas A. Wynn. “Protective and pathogenic functions of macrophage subsets”. In: **Nature Reviews Immunology** 11.11 (2011), pp. 723–737. ISSN: 1474-1733. DOI: 10.1038/nri3073.

[20] Ekaterina Nikitina et al. “Monocytes and Macrophages as Viral Targets and Reservoirs”. In: **International Journal of Molecular Sciences** 19.9 (2018), p. 2821. DOI: 10.3390/ijms19092821.

[21] Georges Herbein and Audrey Varin. “The macrophage in HIV-1 infection: From activation to deactivation?” In: **Retrovirology** 7.1 (2010), p. 33. DOI: 10.1186/1742-4690-7-33.

[22] Claudia R. Avalos et al. “Brain Macrophages in Simian Immunodeficiency Virus-Infected, Antiretroviral-Suppressed Macaques: a Functional Latent Reservoir”. In: **mBio** 8.4 (2017), e01186–17. ISSN: 2161-2129. DOI: 10.1128/mbio.01186-17.

[23] Shu-Wen Wan et al. “The monocyte-macrophage-mast cell axis in dengue pathogenesis”. In: **Journal of Biomedical Science** 25.1 (2018), p. 77. ISSN: 1021-7770. DOI: 10.1186/s12929-018-0482-9.

[24] Jacky Flipse et al. “Antibody-Dependent Enhancement of Dengue Virus Infection in Primary Human Macrophages; Balancing Higher Fusion against Antiviral Responses”. In: **Scientific Reports** 6.1 (2016), p. 29201. DOI: 10.1038/srep29201.

[25] Naoko Uno and Ted M. Ross. “Dengue virus and the host innate immune response”. In: **Emerging Microbes & Infections** 7.1 (2018), p. 167. DOI: 10.1038/s41426-018-0168-0.

[26] Troy D. Cline, Donald Beck, and Elizabeth Bianchini. “Influenza virus replication in macrophages: balancing protection and pathogenesis”. In: **Journal of General Virology** 98.10 (2017), pp. 2401–2412. ISSN: 0022-1317. DOI: 10.1099/jgv.0.000922.

[27] Sukhmani Bedi et al. “A Defect in Influenza A Virus Particle Assembly Specific to Primary Human Macrophages”. In: **mBio** 9.5 (2018), e01916–18. ISSN: 2161-2129. DOI: 10.1128/mbio.01916-18.

[28] Marieke Hoeve et al. “Influenza Virus A Infection of Human Monocyte and Macrophage Subpopulations Reveals Increased Susceptibility Associated with Cell Differentiation”. In: **PLoS ONE** 7.1 (2012), e29443. DOI: 10.1371/journal.pone.0029443.

[29] Michelle D Tate et al. “Critical Role of Airway Macrophages in Modulating Disease Severity during Influenza Virus Infection of Mice”. In: **Journal of Virology** 84.15 (2010), pp. 7569–7580. ISSN: 0022-538X. DOI: 10.1128/jvi.00291-10.

[30] Urška Repnik, Miomir Knezevic, and Matjaz Jeras. “Simple and cost-effective isolation of monocytes from buffy coats”. In: **Journal of Immunological Methods** 278.1-2 (2003), pp. 283–292. ISSN: 0022-1759. DOI: 10.1016/s0022-1759(03)00231-x.

[31] Marlene C. Nielsen, Morten N. Andersen, and Holger J. Møller. “Monocyte isolation techniques significantly impact the phenotype of both isolated monocytes and derived

macrophages in vitro". In: **Immunology** 159.1 (2020), pp. 63–74. ISSN: 0019-2805. DOI: 10.1111/imm.13125.

[32] Wasaporn Chanput, Vera Peters, and Harry Wichers. **The Impact of Food Bioactives on Health, in vitro and ex vivo models**. Springer, 2015, pp. 147–159. DOI: 10.1007/978-3-319-16104-4_14.

[33] Wasaporn Chanput, Jurriaan J. Mes, and Harry J. Wichers. "THP-1 cell line: An in vitro cell model for immune modulation approach". In: **International Immunopharmacology** 23.1 (2014), pp. 37–45. ISSN: 1567-5769. DOI: 10.1016/j.intimp.2014.08.002.

[34] Edana Cassol et al. "Monocyte-derived macrophages and myeloid cell lines as targets of HIV-1 replication and persistence". In: **Journal of Leukocyte Biology** 80.5 (2006), pp. 1018–1030. ISSN: 0741-5400. DOI: 10.1189/jlb.0306150.

[35] Darren M. Riddy et al. "Comparative genotypic and phenotypic analysis of human peripheral blood monocytes and surrogate monocyte-like cell lines commonly used in metabolic disease research". In: **PLoS ONE** 13.5 (2018), e0197177. DOI: 10.1371/journal.pone.0197177.

[36] Xiaosong Liu et al. "Yamanaka factors critically regulate the developmental signaling network in mouse embryonic stem cells". In: **Cell Research** 18.12 (2008), pp. 1177–1189. ISSN: 1001-0602. DOI: 10.1038/cr.2008.309.

[37] Kazutoshi Takahashi and Shinya Yamanaka. "Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors". In: **Cell** 126.4 (2006), pp. 663–676. ISSN: 0092-8674. DOI: 10.1016/j.cell.2006.07.024.

[38] Arun Sharma et al. "Multi-lineage Human iPSC-Derived Platforms for Disease Modeling and Drug Discovery". In: **Cell Stem Cell** 26.3 (2020), pp. 309–329. ISSN: 1934-5909. DOI: 10.1016/j.stem.2020.02.011.

[39] Bonnie van Wilgenburg et al. “Efficient, Long Term Production of Monocyte-Derived Macrophages from Human Pluripotent Stem Cells under Partly-Defined and Fully-Defined Conditions”. In: **PLoS ONE** 8.8 (2013), e71098. DOI: 10.1371/journal.pone.0071098.

[40] Di Cui et al. “High-Yield Human Induced Pluripotent Stem Cell-Derived Monocytes and Macrophages Are Functionally Comparable With Primary Cells”. In: **Frontiers in Cell and Developmental Biology** 9 (2021), p. 656867. ISSN: 2296-634X. DOI: 10.3389/fcell.2021.656867.

[41] Angela M Caliendo et al. “Better Tests, Better Care: Improved Diagnostics for Infectious Diseases”. In: **Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America** 57.Suppl 3 (2013), S139–S170. ISSN: 1058-4838. DOI: 10.1093/cid/cit578.

[42] Gregory A. Storch. “Diagnostic Virology”. In: **Clinical Infectious Diseases** 31.3 (2000), pp. 739–751. ISSN: 1058-4838. DOI: 10.1086/314015.

[43] Alexander Krüttgen et al. “Comparison of the SARS-CoV-2 Rapid Antigen Test to the Real Star Sars-CoV-2 RT PCR Kit”. In: **Journal of Virological Methods** 288 (2020), p. 114024. ISSN: 0166-0934. DOI: 10.1016/j.jviromet.2020.114024.

[44] Jonathan J Deeks et al. “Antibody tests for identification of current and past infection with SARS-CoV-2”. In: **The Cochrane Database of Systematic Reviews** 2020.6 (2020), p. CD013652. ISSN: 1465-1858. DOI: 10.1002/14651858.cd013652.

[45] Chantal B F Vogels et al. “SalivaDirect: A simplified and flexible platform to enhance SARS-CoV-2 testing capacity”. In: **Med** (2020). ISSN: 2666-6340. DOI: 10.1016/j.medj.2020.12.010.

[46] Qing Yang et al. “Saliva TwoStep for rapid detection of asymptomatic SARS-CoV-2 carriers”. In: **eLife** 10 (2021), e65113. DOI: 10.7554/elife.65113.

[47] Steph H Tan et al. “Saliva as a gold-standard sample for SARS-CoV-2 detection”. In: **The Lancet. Respiratory Medicine** 9.6 (2021), pp. 562–564. ISSN: 2213-2600. DOI: 10.1016/s2213-2600(21)00178-8.

[48] Getachew Tesfaye Beyene et al. “Saliva is superior over nasopharyngeal swab for detecting SARS-CoV2 in COVID-19 patients”. In: **Scientific Reports** 11.1 (2021), p. 22640. DOI: 10.1038/s41598-021-02097-2.

[49] Kelvin Kai-Wang To et al. “Consistent detection of 2019 novel coronavirus in saliva.” In: **Clinical infectious diseases : an official publication of the Infectious Diseases Society of America** (2020). ISSN: 1058-4838. DOI: 10.1093/cid/ciaa149.

[50] Wei-Kung Wang et al. “Detection of SARS-associated Coronavirus in Throat Wash and Saliva in Early Diagnosis - Volume 10, Number 7—July 2004 - Emerging Infectious Diseases journal - CDC”. In: **Emerging Infectious Diseases** 10.7 (2004), pp. 1213–1219. ISSN: 1080-6040. DOI: 10.3201/eid1007.031113.

[51] Jung Yoon et al. “The use of saliva specimens for detection of influenza A and B viruses by rapid influenza diagnostic tests”. In: **Journal of Virological Methods** 243 (2017), pp. 15–19. ISSN: 0166-0934. DOI: 10.1016/j.jviromet.2017.01.013.

[52] Joan L. Robinson et al. “Use of Throat Swab or Saliva Specimens for Detection of Respiratory Viruses in Children”. In: **Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America** 46.7 (2008), e61–e64. ISSN: 1058-4838. DOI: 10.1086/529386.

[53] Young-gon Kim et al. “Comparison between Saliva and Nasopharyngeal Swab Specimens for Detection of Respiratory Viruses by Multiplex Reverse Transcription-PCR”. In: **Journal of Clinical Microbiology** 55.1 (2017), pp. 226–233. ISSN: 0095-1137. DOI: 10.1128/jcm.01704-16.

[54] Anna Ritah Namuganga et al. "Suitability of saliva for Tuberculosis diagnosis: comparing with serum". In: **BMC Infectious Diseases** 17.1 (2017), p. 600. DOI: 10.1186/s12879-017-2687-z.

[55] Amy L. Leber et al. "Multicenter Evaluation of BioFire FilmArray Respiratory Panel 2 for Detection of Viruses and Bacteria in Nasopharyngeal Swab Samples". In: **Journal of Clinical Microbiology** 56.6 (2018), e01945–17. ISSN: 0095-1137. DOI: 10.1128/jcm.01945-17.

[56] Sze Hwei Lee et al. "Performance of a multiplex PCR pneumonia panel for the identification of respiratory pathogens and the main determinants of resistance from the lower respiratory tract specimens of adult patients in intensive care units". In: **Journal of Microbiology, Immunology and Infection** 52.6 (2019), pp. 920–928. ISSN: 1684-1182. DOI: 10.1016/j.jmii.2019.10.009.

[57] Caitlin N. Murphy et al. "Multicenter Evaluation of the BioFire FilmArray Pneumonia/Pneumonia Plus Panel for Detection and Quantification of Agents of Lower Respiratory Tract Infection". In: **Journal of Clinical Microbiology** 58.7 (2020), e00128–20. ISSN: 0095-1137. DOI: 10.1128/jcm.00128-20.

[58] Telma R Poloni et al. "Detection of dengue virus in saliva and urine by real time RT-PCR". In: **Virology Journal** 7.1 (2010), pp. 22–22. DOI: 10.1186/1743-422x-7-22.

[59] Giuseppina Liuzzi et al. "Analysis of HIV-1 load in blood, semen and saliva: evidence for different viral compartments in a cross-sectional and longitudinal study". In: **AIDS** 10.14 (1996), F51–F56. ISSN: 0269-9370. DOI: 10.1097/00002030-199612000-00001.

[60] Pauline Vetter et al. "Ebola Virus Shedding and Transmission: Review of Current Evidence". In: **The Journal of Infectious Diseases** 214.suppl_3 (2016), S177–S184. ISSN: 0022-1899. DOI: 10.1093/infdis/jiw254.

[61] Weihong Yan et al. “Systematic comparison of the human saliva and plasma proteomes”. In: **PROTEOMICS - Clinical Applications** 3.1 (2009), pp. 116–134. ISSN: 1862-8354. DOI: 10.1002/prca.200800140.

[62] Hajer Jasim et al. “Saliva as a medium to detect and measure biomarkers related to pain”. In: **Scientific Reports** 8.1 (2018), p. 3220. DOI: 10.1038/s41598-018-21131-4.

[63] Daniel Malamud. “Saliva as a Diagnostic Fluid”. In: **Dental Clinics of North America** 55.1 (2011), pp. 159–178. ISSN: 0011-8532. DOI: 10.1016/j.cden.2010.08.004.

[64] Xiaowei Song et al. “Oral squamous cell carcinoma diagnosed from saliva metabolic profiling”. In: **Proceedings of the National Academy of Sciences** 117.28 (2020), pp. 16167–16173. ISSN: 0027-8424. DOI: 10.1073/pnas.2001395117.

[65] Shigeo Ishikawa et al. “Identification of salivary metabolomic biomarkers for oral cancer screening”. In: **Scientific Reports** 6.1 (2016), p. 31520. DOI: 10.1038/srep31520.

[66] Stuart S. Kassan and Haralampos M. Moutsopoulos. “Clinical Manifestations and Early Diagnosis of Sjögren Syndrome”. In: **Archives of Internal Medicine** 164.12 (2004), pp. 1275–1284. ISSN: 0003-9926. DOI: 10.1001/archinte.164.12.1275.

[67] Zachary E Holcomb et al. “Host-Based Peripheral Blood Gene Expression Analysis for Diagnosis of Infectious Diseases”. In: **Journal of Clinical Microbiology** 55.2 (2017), pp. 360–368. ISSN: 0095-1137. DOI: 10.1128/jcm.01057-16.

[68] Matthias Eberl et al. “Pathogen-Specific Immune Fingerprints during Acute Infection: The Diagnostic Potential of Human T-Cells”. In: **Frontiers in Immunology** 5 (2014), p. 572. ISSN: 1664-3224. DOI: 10.3389/fimmu.2014.00572.

[69] Yudong D. He et al. "The Optimization and Biological Significance of a 29-Host-Immune-mRNA Panel for the Diagnosis of Acute Infections and Sepsis". In: **Journal of Personalized Medicine** 11.8 (2021), p. 735. ISSN: 2075-4426. DOI: 10.3390/jpm11080735.

[70] Julie G. Burel et al. "Host Transcriptomics as a Tool to Identify Diagnostic and Mechanistic Immune Signatures of Tuberculosis". In: **Frontiers in Immunology** 10 (2019), p. 221. DOI: 10.3389/fimmu.2019.00221.

[71] Samuel Channon-Wells and Daniel O'Connor. "Host gene signature shows promise to distinguish bacterial and viral infections". In: **The Lancet. Digital Health** 3.8 (2021), e465–e466. ISSN: 2589-7500. DOI: 10.1016/s2589-7500(21)00136-9.

[72] Octavio Ramilo and Asunción Mejías. "Shifting the Paradigm: Host Gene Signatures for Diagnosis of Infectious Diseases". In: **Cell Host & Microbe** 6.3 (2009), pp. 199–200. ISSN: 1931-3128. DOI: 10.1016/j.chom.2009.08.007.

[73] Andrew R. Moore et al. "Prospective validation of an 11-gene mRNA host response score for mortality risk stratification in the intensive care unit". In: **Scientific Reports** 11.1 (2021), p. 13062. DOI: 10.1038/s41598-021-91201-7.

[74] Simone A. Thair et al. "Transcriptomic similarities and differences in host response between SARS-CoV-2 and other viral infections". In: **iScience** 24.1 (2020), p. 101947. ISSN: 2589-0042. DOI: 10.1016/j.isci.2020.101947.

[75] World Health Organization. **Novel Coronavirus(2019-nCoV) Situation Report**. Jan. 2020. URL: https://apps.who.int/iris/bitstream/handle/10665/330760/nCoVsitrep21Jan2020-eng.pdf?sequence=3%5C&isAllowed=y.

[76] Peng Zhou et al. "A pneumonia outbreak associated with a new coronavirus of probable bat origin." In: **Nature** 579.7798 (2020), pp. 270–273. ISSN: 0028-0836. DOI: 10.1038/s41586-020-2012-7.

[77] Tommy Tsan-Yuk Lam et al. "Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins". In: **Nature** (2020), pp. 1–6. ISSN: 0028-0836. DOI: 10.1038/s41586-020-2169-0.

[78] Lea A Nikolai et al. "Asymptomatic SARS Coronavirus 2 infection: Invisible yet invincible". In: **International Journal of Infectious Diseases** 100 (2020), pp. 112–116. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2020.08.076.

[79] Piero Poletti et al. "Probability of symptoms and critical disease after SARS-CoV-2 infection". In: **arXiv** (2020). eprint: 2006.08471.

[80] Daniel C Payne et al. "SARS-CoV-2 Infections and Serologic Responses from a Sample of U.S. Navy Service Members — USS Theodore Roosevelt, April 2020". In: **Morbidity and Mortality Weekly Report** 69.23 (2020), pp. 714–721. ISSN: 0149-2195. DOI: 10.15585/mmwr.mm6923e4.

[81] Xingfei Pan et al. "Asymptomatic cases in a family cluster with SARS-CoV-2 infection". In: **The Lancet Infectious Diseases** 20.4 (2020), pp. 410–411. ISSN: 1473-3099. DOI: 10.1016/s1473-3099(20)30114-6.

[82] Xiquan Yan et al. "Duration of SARS-CoV-2 viral RNA in asymptomatic carriers". In: **Critical Care** 24.1 (2020), p. 245. DOI: 10.1186/s13054-020-02952-0.

[83] Yan Bai et al. "Presumed Asymptomatic Carrier Transmission of COVID-19". In: **JAMA** 323.14 (2020), pp. 1406–1407. ISSN: 0098-7484. DOI: 10.1001/jama.2020.2565.

[84] Feng Ye et al. "Delivery of infection from asymptomatic carriers of COVID-19 in a familial cluster". In: **International Journal of Infectious Diseases** 94 (2020), pp. 133–138. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2020.03.042.

[85] Alvin J Ing, Christine Cocks, and Jeffery Peter Green. "COVID-19: in the footsteps of Ernest Shackleton". In: **Thorax** 75.8 (2020), pp. 693–694. ISSN: 0040-6376. DOI: 10.1136/thoraxjnl-2020-215091.

[86] Zhang Yongchen et al. "Different longitudinal patterns of nucleic acid and serology testing results based on disease severity of COVID-19 patients". In: **Emerging Microbes & Infections** 9.1 (2020), pp. 1–14. DOI: 10.1080/22221751.2020.1756699.

[87] Akira Endo et al. "Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China". In: **Wellcome Open Research** 5 (2020), p. 67. ISSN: 2398-502X. DOI: 10.12688/wellcomeopenres.15842.3.

[88] Dillon C. Adam et al. "Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong". In: **Nature Medicine** 26.11 (2020), pp. 1714–1719. ISSN: 1078-8956. DOI: 10.1038/s41591-020-1092-0.

[89] Kai Kupferschmidt. "Why do some COVID-19 patients infect many others, whereas most don't spread the virus at all?" In: **Science** (2020). ISSN: 0036-8075.

[90] Qifang Bi et al. "Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study". In: **The Lancet Infectious Diseases** 20.8 (2020), pp. 911–919. ISSN: 1473-3099. DOI: 10.1016/s1473-3099(20)30287-5.

[91] Anne L. Wyllie et al. "Saliva or Nasopharyngeal Swab Specimens for Detection of SARS-CoV-2". In: **New England Journal of Medicine** 383.13 (2020), pp. 1283–1286. ISSN: 0028-4793. DOI: 10.1056/nejmc2016359.

[92] Julio Silva et al. "Saliva viral load is a dynamic unifying correlate of COVID-19 severity and mortality". In: **medRxiv : the preprint server for health sciences** (2021). DOI: 10.1101/2021.01.04.21249236. URL: https://www.medrxiv.org/content/10.1101/2020.07.16.20150250v2.full.pdf.

[93] Erik Volz et al. "Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data". In: **medRxiv** (2021). DOI: 10.1101/2020.12.30.20249034.

[94] Diana Rose E. Ranoa et al. "Saliva-Based Molecular Testing for SARS-CoV-2 that Bypasses RNA Extraction". In: **bioRxiv** (2020), p. 2020.06.18.159434. DOI: 10.1101/2020.06.18.159434. eprint: https://www.biorxiv.org/content/early/2020/06/18/2020.06.18.159434.full.pdf.

[95] Sang Hyun Ra et al. "Upper respiratory viral load in asymptomatic individuals and mildly symptomatic patients with SARS-CoV-2 infection". In: **Thorax** 76.1 (2021), pp. 61–63. ISSN: 0040-6376. DOI: 10.1136/thoraxjnl-2020-215042.

[96] Niall J Lennon et al. "Comparison of viral levels in individuals with or without symptoms at time of COVID-19 testing among 32,480 residents and staff of nursing homes and assisted living facilities in Massachusetts." In: **medRxiv** (2020). DOI: 10.1101/2020.07.20.20157792.

[97] Tongyang Xiao et al. "Early viral clearance and antibody kinetics of COVID-19 among asymptomatic carriers". In: **medRxiv** (2020). DOI: 10.1101/2020.04.28.20083139.

[98] Nguyen Van Vinh Chau et al. "The natural history and transmission potential of asymptomatic SARS-CoV-2 infection". In: **Clinical Infectious Diseases** 71.10 (2020), ciaa711–. ISSN: 1537-6591. DOI: 10.1093/cid/ciaa711.

[99] Melissa M Arons et al. "Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility". In: **New England Journal of Medicine** (2020). ISSN: 0028-4793. DOI: 10.1056/nejmoa2008457.

[100] Muge Cevik et al. "SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis".

In: **The Lancet Microbe** 2.1 (2021), e13–e22. ISSN: 2666-5247. DOI: 10.1016/s2666-5247(20)30172-5.

[101] Chung-Guei Huang et al. "Culture-Based Virus Isolation To Evaluate Potential Infectivity of Clinical Specimens Tested for COVID-19". In: **Journal of Clinical Microbiology** 58.8 (2020). ISSN: 0095-1137. DOI: 10.1128/jcm.01068-20.

[102] Roman Wölfel et al. "Virological assessment of hospitalized patients with COVID-2019". In: **Nature** 581.7809 (2020), pp. 465–469. ISSN: 0028-0836. DOI: 10.1038/s41586-020-2196-x.

[103] Bernard La Scola et al. "Viral RNA load as determined by cell culture as a management tool for discharge of SARS-CoV-2 patients from infectious disease wards". In: **European Journal of Clinical Microbiology & Infectious Diseases** 39.6 (2020), pp. 1059–1061. ISSN: 0934-9723. DOI: 10.1007/s10096-020-03913-9.

[104] Rita Jaafar et al. "Correlation Between 3790 Quantitative Polymerase Chain Reaction–Positives Samples and Positive Cell Cultures, Including 1941 Severe Acute Respiratory Syndrome Coronavirus 2 Isolates". In: **Clinical Infectious Disease** (2020).

[105] Jared Bullard et al. "Predicting infectious SARS-CoV-2 from diagnostic samples". In: **Clinical Infectious Diseases** 71.10 (2020), ciaa638–. ISSN: 1537-6591. DOI: 10.1093/cid/ciaa638.

[106] Min-Chui Kim et al. "Duration of Culturable SARS-CoV-2 in Hospitalized Patients with Covid-19". In: **New England Journal of Medicine** 384.7 (2021), pp. 669–671.

[107] Garrett A Perchetti et al. "Analytical Sensitivity of the Abbott BinaxNOW COVID-19 Ag CARD." In: **Journal of clinical microbiology** (2020). DOI: 10.1128/jcm.02880-20.

[108] U.S. Food & Drug Administration. **SARS-CoV-2 Reference Panel Comparative Data**. 2020. URL: https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/sars-cov-2-reference-panel-comparative-data (visited on 04/07/2021).

[109] Xi He et al. "Temporal dynamics in viral shedding and transmissibility of COVID-19". In: **Nature Medicine** 26.5 (2020), pp. 672–675. ISSN: 1078-8956. DOI: 10.1038/s41591-020-0869-5.

[110] Michael Marks et al. "Transmission of COVID-19 in 282 clusters in Catalonia, Spain: a cohort study". In: **The Lancet Infectious Diseases** (2021). ISSN: 1473-3099. DOI: 10.1016/s1473-3099(20)30985-3.

[111] Kristen K Bjorkman et al. "Higher Viral Load Drives Infrequent Severe Acute Respiratory Syndrome Coronavirus 2 Transmission Between Asymptomatic Residence Hall Roommates". In: **The Journal of Infectious Diseases** 224.8 (July 2021), pp. 1316–1324. ISSN: 0022-1899. DOI: 10.1093/infdis/jiab386. eprint: https://academic.oup.com/jid/article-pdf/224/8/1316/40978920/jiab386.pdf. URL: https://doi.org/10.1093/infdis/jiab386.

[112] Kanecia O Zimmerman et al. "Incidence and Secondary Transmission of SARS-CoV-2 Infections in Schools". In: **Pediatrics** (2021), e2020048090. ISSN: 0031-4005. DOI: 10.1542/peds.2020-048090.

[113] Greg Gibson et al. "Surveillance-to-Diagnostic Testing Program for Asymptomatic SARS-CoV-2 Infections on a Large, Urban Campus - Georgia Institute of Technology, Fall 2020". In: **medRxiv** 384.7 (2021), pp. 669–671. DOI: 10.1101/2021.01.28.21250700.

[114] Zachary J. Madewell et al. "Household Transmission of SARS-CoV-2". In: **JAMA Network Open** 3.12 (2020), e2031756. DOI: 10.1001/jamanetworkopen.2020.31756.

[115] Stephen M Kissler et al. "SARS-CoV-2 viral dynamics in acute infections". In: **medRxiv** (2020). DOI: 10.1101/2020.10.21.20217042.

[116] Eu Suk Kim et al. "Clinical Course and Outcomes of Patients with Severe Acute Respiratory Syndrome Coronavirus 2 Infection: a Preliminary Report of the First 28 Patients from the Korean Cohort Study on COVID-19". In: **Journal of Korean Medical Science** 35.13 (2020), e142. ISSN: 1011-8934. DOI: 10.3346/jkms.2020.35.e142.

[117] Lirong Zou et al. "SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients". In: **New England Journal of Medicine** 382.12 (2020), pp. 1177–1179. ISSN: 0028-4793. DOI: 10.1056/nejmc2001737.

[118] Yuan Hou et al. "New insights into genetic susceptibility of COVID-19: an ACE2 and TMPRSS2 polymorphism analysis". In: **BMC Medicine** 18.1 (2020), p. 216. DOI: 10.1186/s12916-020-01673-z.

[119] Elisa Benetti et al. "ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population". In: **European Journal of Human Genetics** 28.11 (2020), pp. 1602–1614. ISSN: 1018-4813. DOI: 10.1038/s41431-020-0691-z.

[120] Jesse Fajnzylber et al. "SARS-CoV-2 viral load is associated with increased disease severity and mortality". In: **Nature Communications** 11.1 (2020), p. 5493. DOI: 10.1038/s41467-020-19057-5.

[121] Michael Kidd et al. "S-variant SARS-CoV-2 is associated with significantly higher viral loads in samples tested by ThermoFisher TaqPath RT-QPCR". In: **medRxiv** (Dec. 2020). DOI: 10.1101/2020.12.24.20248834.

[122] Larremore et al. "Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening". In: **Science Advances** (2020).

[123] Kelvin Kai-Wang To et al. "Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study". In: **The Lancet Infectious Diseases** 20.5 (2020), pp. 565–574. ISSN: 1473-3099. DOI: 10.1016/s1473-3099(20)30196-1.

[124] Wei Zhang et al. "Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes". In: **Emerging Microbes & Infections** 9.1 (2020), pp. 386–389. DOI: 10.1080/22221751.2020.1729071.

[125] Fatih M. Hanege et al. "SARS-CoV-2 Presence in the Saliva, Tears, and Cerumen of COVID-19 Patients". In: **The Laryngoscope** (2020). ISSN: 0023-852X. DOI: 10.1002/lary.29218.

[126] Gary W Procop et al. "A Direct Comparison of Enhanced Saliva to Nasopharyngeal Swab for the Detection of SARS-CoV-2 in Symptomatic Patients". In: **Journal of Clinical Microbiology** 58.11 (2020). ISSN: 0095-1137. DOI: 10.1128/jcm.01946-20.

[127] Jin Gu Yoon et al. "Clinical Significance of a High SARS-CoV-2 Viral Load in the Saliva". In: **Journal of Korean Medical Science** 35.20 (2020), e195. ISSN: 1011-8934. DOI: 10.3346/jkms.2020.35.e195.

[128] Isao Yokota et al. "Equivalent SARS-CoV-2 viral loads by PCR between nasopharyngeal swab and saliva in symptomatic patients". In: **Scientific Reports** 11.1 (2021), p. 4500. DOI: 10.1038/s41598-021-84059-2.

[129] Jialou Zhu et al. "Viral dynamics of SARS-CoV-2 in saliva from infected patients". In: **Journal of Infection** 81.3 (2020), e48–e50. ISSN: 0163-4453. DOI: 10.1016/j.jinf.2020.06.059.

[130] William F Wright and Philip A Mackowiak. "Why Temperature Screening for Coronavirus Disease 2019 With Noncontact Infrared Thermometers Does Not Work".

In: **Open Forum Infectious Diseases** 8.1 (2020), ofaa603. ISSN: 2328-8957. DOI: 10.1093/ofid/ofaa603.

[131] Tsugunori Notomi et al. "Loop-mediated isothermal amplification of DNA". In: **Nucleic Acids Research** 28.12 (2000), e63–e63. ISSN: 0305-1048. DOI: 10.1093/nar/28.12.e63.

[132] Connie D Brewster et al. "LAMP assays of Zika virus and other infectious agents will inevitably see expanded use due to their simplicity, sensitivity, specificity, and economy". In: **Annals of Translational Medicine** 6.10 (2018), pp. 196–196. ISSN: 2305-5847. DOI: 10.21037/18871.

[133] Nunya Chotiwan et al. "Rapid and specific detection of Asian- and African-lineage Zika viruses". In: **Science Translational Medicine** 9.388 (2017), eaag0538. ISSN: 1946-6234. DOI: 10.1126/scitranslmed.aag0538.

[134] Max J Kellner et al. "A rapid, highly sensitive and open-access SARS-CoV-2 detection assay for laboratory and home testing". In: **bioRxiv** (2020), p. 2020.06.23.166397. DOI: 10.1101/2020.06.23.166397.

[135] Nathan A. Tanner, Yinhua Zhang, and Thomas C. Evans. "Visual detection of isothermal nucleic acid amplification using pH-sensitive dyes". In: **BioTechniques** 58.2 (2015), pp. 59–68. ISSN: 0736-6205. DOI: 10.2144/000114253.

[136] Patrick Hardinge and James A. H. Murray. "Reduced False Positives and Improved Reporting of Loop-Mediated Isothermal Amplification using Quenched Fluorescent Primers". In: **Scientific Reports** 9.1 (2019), p. 7400. DOI: 10.1038/s41598-019-43817-z.

[137] Gregoire Seyrig et al. "Selection of fluorescent DNA dyes for real-time LAMP with portable and simple optics". In: **Journal of Microbiological Methods** 119 (2015), pp. 223–227. ISSN: 0167-7012. DOI: 10.1016/j.mimet.2015.11.004.

[138] Yasuyoshi Mori et al. "Detection of Loop-Mediated Isothermal Amplification Reaction by Turbidity Derived from Magnesium Pyrophosphate Formation". In: **Biochemical and Biophysical Research Communications** 289.1 (2001), pp. 150–154. ISSN: 0006-291X. DOI: 10.1006/bbrc.2001.5921.

[139] Pavana Khan, Lauren M Aufdembrink, and Aaron E Engelhart. "Isothermal SARS-CoV-2 Diagnostics: Tools for Enabling Distributed Pandemic Testing as a Means of Supporting Safe Reopenings". In: **ACS Synthetic Biology** 9.11 (2020), pp. 2861–2880. ISSN: 2161-5063. DOI: 10.1021/acssynbio.0c00359.

[140] Guillaume Butler-Laporte et al. "Comparison of Saliva and Nasopharyngeal Swab Nucleic Acid Amplification Testing for Detection of SARS-CoV-2". In: **JAMA Internal Medicine** 181.3 (2021). ISSN: 2168-6106. DOI: 10.1001/jamainternmed.2020.8876.

[141] Isao Yokota et al. "Mass screening of asymptomatic persons for SARS-CoV-2 using saliva". In: **Clinical Infectious Diseases** (2020), p. 1388. ISSN: 1058-4838. DOI: 10.1093/cid/ciaa1388.

[142] Sanchita Bhadra et al. "High-surety isothermal amplification and detection of SARS-CoV-2, including with crude enzymes". In: **bioRxiv** (2020), p. 2020.04.13.039941. DOI: 10.1101/2020.04.13.039941.

[143] Michael J Flynn et al. "A simple direct RT-LAMP SARS-CoV-2 saliva diagnostic". In: **medRxiv** (2020). DOI: 10.1101/2020.11.19.20234948.

[144] Matthew A Lalli et al. "Rapid and extraction-free detection of SARS-CoV-2 from saliva with colorimetric LAMP". In: **medRxiv : the preprint server for health sciences** (2020). DOI: 10.1101/2020.05.07.20093542.

[145] Laura E Lamb et al. "Rapid detection of novel coronavirus/Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) by reverse transcription-loop-mediated

isothermal amplification". In: **PLOS ONE** 15.6 (2020), e0234682. DOI: 10.1371/journal.pone.0234682.

[146] Mayu Nagura-Ikeda et al. "Clinical Evaluation of Self-Collected Saliva by Quantitative Reverse Transcription PCR (RT-qPCR), Direct RT-qPCR, Reverse Transcription Loop Mediated Isothermal Amplification, and a Rapid Antigen Test To Diagnose COVID-19". In: **Journal of Clinical Microbiology** 58.9 (2020). ISSN: 0095-1137. DOI: 10.1128/jcm.01438-20.

[147] Brian A. Rabe and Constance Cepko. "SARS-CoV-2 detection using isothermal amplification and a rapid, inexpensive protocol for sample inactivation and purification". In: **Proceedings of the National Academy of Sciences** 117.39 (2020), pp. 24450–24458. ISSN: 0027-8424. DOI: 10.1073/pnas.2011221117.

[148] Keisuke Taki et al. "SARS-CoV-2 detection by fluorescence loop-mediated isothermal amplification with and without RNA extraction". In: **Journal of Infection and Chemotherapy** 27.2 (2020), pp. 410–412. ISSN: 1341-321X. DOI: 10.1016/j.jiac.2020.10.029.

[149] Simon J. S. Cameron et al. "The human salivary microbiome exhibits temporal stability in bacterial diversity". In: **FEMS Microbiology Ecology** 91.9 (2015), fiv091. ISSN: 0168-6496. DOI: 10.1093/femsec/fiv091.

[150] Kelly A. Curtis et al. "A Multiplexed RT-LAMP Assay for Detection of Group M HIV-1 in Plasma or Whole Blood". In: **Journal of Virological Methods** 255 (2018), pp. 91–97. ISSN: 0166-0934. DOI: 10.1016/j.jviromet.2018.02.012.

[151] Christophe Batéjat et al. "Heat inactivation of the Severe Acute Respiratory Syndrome Coronavirus 2". In: **Journal of Biosafety and Biosecurity** 3.1 (2021), pp. 1–3. ISSN: 2588-9338. DOI: 10.1016/j.jobb.2020.12.001.

[152] Dan Davidi et al. “Amplicon contamination in labs masquerades as COVID19 in surveillance tests”. In: **medRxiv** (2020). DOI: 10.1101/2020.12.08.20244525.

[153] Yi Tang, Hao Chen, and Youxiang Diao. “Advanced uracil DNA glycosylase supplemented realtime reverse transcription loop mediated isothermal amplification (UDG-rRT-LAMP) method for universal and specific detection of Tembusu virus”. In: **Scientific Reports** 6.1 (2016), p. 27605. DOI: 10.1038/srep27605.

[154] Ramona Mögling et al. “Delayed Laboratory Response to COVID-19 Caused by Molecular Diagnostic Contamination - Volume 26, Number 8—August 2020 - Emerging Infectious Diseases journal - CDC”. In: **Emerging Infectious Diseases** 26.8 (2020), pp. 1944–1946. ISSN: 1080-6040. DOI: 10.3201/eid2608.201843.

[155] Viet Loan Dao Thi et al. “A colorimetric RT-LAMP assay and LAMP-sequencing for detecting SARS-CoV-2 RNA in clinical samples”. In: **Science Translational Medicine** 12.556 (2020), eabc7075. ISSN: 1946-6234. DOI: 10.1126/scitranslmed.abc7075.

[156] Micah T McClain et al. “A blood-based host gene expression assay for early detection of respiratory viral infection: an index-cluster prospective cohort study”. In: **The Lancet Infectious Diseases** 21.3 (2021), pp. 396–404. ISSN: 1473-3099. DOI: 10.1016/s1473-3099(20)30486-2.

[157] Rishi K Gupta 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”. In: **The Lancet Microbe** 2.10 (2021), e508–e517. ISSN: 2666-5247. DOI: 10.1016/s2666-5247(21)00146-4.

[158] Russell C. Ault et al. “Blood RNA signatures predict recent tuberculosis exposure in mice, macaques and humans”. In: **Scientific Reports** 10.1 (2020), p. 16873. DOI: 10.1038/s41598-020-73942-z.

[159] Ephraim L Tsalik et al. "Discriminating Bacterial and Viral Infection Using a Rapid Host Gene Expression Test*". In: **Critical Care Medicine** 49.10 (2021), pp. 1651–1663. ISSN: 0090-3493. DOI: 10.1097/ccm.0000000000005085.

[160] Aimee K. Zaas et al. "Blood Gene Expression Signatures Predict Invasive Candidiasis". In: **Science Translational Medicine** 2.21 (2010), 21ra17–21ra17. ISSN: 1946-6234. DOI: 10.1126/scitranslmed.3000715.

[161] Leo McHugh et al. "A Molecular Host Response Assay to Discriminate Between Sepsis and Infection-Negative Systemic Inflammation in Critically Ill Patients: Discovery and Validation in Independent Cohorts". In: **PLoS Medicine** 12.12 (2015), e1001916. ISSN: 1549-1277. DOI: 10.1371/journal.pmed.1001916.

[162] Santtu Heinonen et al. "Rhinovirus Detection in Symptomatic and Asymptomatic Children: Value of Host Transcriptome Analysis". In: **American Journal of Respiratory and Critical Care Medicine** 193.7 (2016), pp. 772–782. ISSN: 1073-449X. DOI: 10.1164/rccm.201504-0749oc.

[163] D. L. Sampson et al. "A Four-Biomarker Blood Signature Discriminates Systemic Inflammation Due to Viral Infection Versus Other Etiologies". In: **Scientific Reports** 7.1 (2017), p. 2914. DOI: 10.1038/s41598-017-02325-8.

[164] Maike Sperk et al. "Utility of Proteomics in Emerging and Re-Emerging Infectious Diseases Caused by RNA Viruses". In: **Journal of Proteome Research** 19.11 (2020), pp. 4259–4274. ISSN: 1535-3893. DOI: 10.1021/acs.jproteome.0c00380.

[165] Arindam Chakraborty and Singh Monica. "Host biomarkers for early diagnosis of infectious diseases: A comprehensive review". In: **International Journal of Clinical Microbiology and Biochemical Technology** 2.1 (2019), pp. 001–007. DOI: 10.29328/journal.ijcmbt.1001005.

[166] Marina Rieder et al. "Serum protein profiling reveals a specific upregulation of the immunomodulatory protein progranulin in COVID-19". In: **The Journal of Infectious Diseases** 223.5 (2020), jiaa741. ISSN: 0022-1899. DOI: 10.1093/infdis/jiaa741.

[167] Thomas W. Burke et al. "Nasopharyngeal Protein Biomarkers of Acute Respiratory Virus Infection". In: **EBioMedicine** 17 (2017), pp. 172–181. ISSN: 2352-3964. DOI: 10.1016/j.ebiom.2017.02.015.

[168] Sandipan Ray et al. "Differential expression of serum/plasma proteins in various infectious diseases: Specific or nonspecific signatures". In: **PROTEOMICS - Clinical Applications** 8.1-2 (2013), pp. 53–72. ISSN: 1862-8346. DOI: 10.1002/prca.201300074.

[169] Anna C. Aschenbrenner et al. "Disease severity-specific neutrophil signatures in blood transcriptomes stratify COVID-19 patients". In: **Genome Medicine** 13.1 (2021), p. 7. DOI: 10.1186/s13073-020-00823-5.

[170] Yijie Zhai et al. "Host Transcriptional Response to Influenza and Other Acute Respiratory Viral Infections – A Prospective Cohort Study". In: **PLOS Pathogens** 11.6 (2015), e1004869. ISSN: 1553-7366. DOI: 10.1371/journal.ppat.1004869.

[171] Qing Yang et al. "Just 2% of SARS-CoV-2-positive individuals carry 90% of the virus circulating in communities". In: **Proceedings of the National Academy of Sciences** 118.21 (2021), e2104547118. ISSN: 0027-8424. DOI: 10.1073/pnas.2104547118.

[172] Daniel S Chertow et al. "Severe Meningoencephalitis in a Case of Ebola Virus Disease: A Case Report". In: **Annals of Internal Medicine** 165.4 (2016), p. 301. ISSN: 0003-4819. DOI: 10.7326/m15-3066.

[173] John C. Kash et al. “Longitudinal peripheral blood transcriptional analysis of a patient with severe Ebola virus disease”. In: **Science Translational Medicine** 9.385 (2017). ISSN: 1946-6234. DOI: 10.1126/scitranslmed.aai9321.

[174] Kathryn Milne et al. “Mapping immune variation and var gene switching in naive hosts infected with Plasmodium falciparum”. In: **eLife** 10 (2021), e62800. DOI: 10.7554/elife.62800.

[175] Emily S. Savela et al. “Quantitative SARS-CoV-2 viral-load curves in paired saliva and nasal swabs inform appropriate respiratory sampling site and analytical test sensitivity required for earliest viral detection”. In: **medRxiv** (2021), p. 2021.04.02.21254771. DOI: 10.1101/2021.04.02.21254771.

[176] Kelly Daniel Sullivan et al. “The COVIDome Explorer researcher portal”. In: **Cell Reports** 36.7 (2021), p. 109527. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2021.109527.

[177] C Lauber et al. “Transcriptome analysis reveals a classical interferon signature induced by IFNlamda4 in human primary cells”. In: **Genes and Immunity** 16.6 (2015), pp. 414–421. ISSN: 1466-4879. DOI: 10.1038/gene.2015.23.

[178] Akul Singhania et al. “A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection”. In: **Nature Communications** 9.1 (June 2018), p. 2308. DOI: 10.1038/s41467-018-04579-w.

[179] Prabhash Kumar Jha et al. “Gene Expression Profiling Reveals the Shared and Distinct Transcriptional Signatures in Human Lung Epithelial Cells Infected With SARS-CoV-2, MERS-CoV, or SARS-CoV: Potential Implications in Cardiovascular Complications of COVID-19”. In: **Frontiers in Cardiovascular Medicine** 7 (2021), p. 623012. ISSN: 2297-055X. DOI: 10.3389/fcvm.2020.623012.

[180] Jeroen L. A. Pennings, Tjeerd G. Kimman, and Riny Janssen. "Identification of a Common Gene Expression Response in Different Lung Inflammatory Diseases in Rodents and Macaques". In: **PLoS ONE** 3.7 (2008), e2596. DOI: 10.1371/journal.pone.0002596.

[181] Aravind Subramanian et al. "A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles". In: **Cell** 171.6 (2017), 1437–1452.e17. ISSN: 0092-8674. DOI: 10.1016/j.cell.2017.10.049.

[182] Weiyi Pan, Qingxuan Wang, and Qianming Chen. "The cytokine network involved in the host immune response to periodontitis". In: **International Journal of Oral Science** 11.3 (2019), p. 30. ISSN: 1674-2818. DOI: 10.1038/s41368-019-0064-z.

[183] Maria Greabu et al. "Saliva–a diagnostic window to the body, both in health and in disease." In: **Journal of medicine and life** 2.2 (2009), pp. 124–32. ISSN: 1844-122X.

[184] Yohann Nédélec et al. "Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens". In: **Cell** 167.3 (2016), 657–669.e21. ISSN: 0092-8674. DOI: 10.1016/j.cell.2016.09.025.

[185] Andrew E Shaw et al. "Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses". English. In: **PLOS Biology** 15.12 (Dec. 2017). Ed. by Harmit Malik, e2004086. ISSN: 1544-9173. DOI: 10.1371/journal.pbio.2004086. URL: http://journals.plos.org/plosbiology/article/file?id=10.1371/journal.pbio.2004086%5C&type=printable.

[186] Brian Bushnell, Jonathan Rood, and Esther Singer. "BBMerge – Accurate paired shotgun read merging via overlap". In: **PLoS ONE** 12.10 (2017), e0185056. DOI: 10.1371/journal.pone.0185056.

[187] Daehwan Kim et al. “Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype”. In: **Nature Biotechnology** 37.8 (2019), pp. 907–915. ISSN: 1087-0156. DOI: 10.1038/s41587-019-0201-4.

[188] Yang Liao, Gordon K. Smyth, and Wei Shi. “The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote”. In: **Nucleic Acids Research** 41.10 (2013), e108–e108. ISSN: 0305-1048. DOI: 10.1093/nar/gkt214.

[189] Tracey Allen K. Freitas et al. “Accurate read-based metagenome characterization using a hierarchical suite of unique signatures”. In: **Nucleic Acids Research** 43.10 (2015), e69–e69. ISSN: 0305-1048. DOI: 10.1093/nar/gkv180.

[190] Davide Risso et al. “Normalization of RNA-seq data using factor analysis of control genes or samples”. In: **Nature Biotechnology** 32.9 (2014), pp. 896–902. ISSN: 1087-0156. DOI: 10.1038/nbt.2931.

[191] Min-Chui Kim et al. “Duration of Culturable SARS-CoV-2 in Hospitalized Patients with Covid-19”. In: **New England Journal of Medicine** 384.7 (2021), pp. 669–671.

[192] David Hare et al. “The Importance of Physiologically Relevant Cell Lines for Studying Virus–Host Interactions”. In: **Viruses** 8.11 (2016), p. 297. DOI: 10.3390/v8110297.

[193] R van Furth, J A Raeburn, and T L van Zwet. “Characteristics of human mononuclear phagocytes.” In: **Blood** 54.2 (1979), pp. 485–500. ISSN: 0006-4971.

[194] Tom P. Monie. **The Innate Immune System, Section 1 - A Snapshot of the Innate Immune System**. Academic Press, 2017. ISBN: 9780128044643. DOI: 10.1016/b978-0-12-804464-3.00001-6.

[195] Kazutoshi Takahashi et al. “Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors”. In: **Cell** 131.5 (2007), pp. 861–872. ISSN: 0092-8674. DOI: 10.1016/j.cell.2007.11.019.

[196] Junying Yu et al. “Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells”. In: **Science** 318.5858 (2007), pp. 1917–1920. ISSN: 0036-8075. DOI: 10.1126/science.1151526.

[197] Ruilin Tian et al. “CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons”. In: **Neuron** 104.2 (2019), 239–255.e12. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2019.07.014.

[198] Duran Surun et al. “Efficient Generation and Correction of Mutations in Human iPS Cells Utilizing mRNAs of CRISPR Base Editors and Prime Editors”. In: **Genes** 11.5 (2020), p. 511. DOI: 10.3390/genes11050511.

[199] J M Wang et al. “Induction of monocyte migration by recombinant macrophage colony-stimulating factor.” In: **The Journal of Immunology** 141.2 (1988), pp. 575–579. ISSN: 0022-1767. URL: https://www.jimmunol.org/content/141/2/575.

[200] Peggie Cheung et al. “Repression of CTSG, ELANE and PRTN3-mediated histone H3 proteolytic cleavage promotes monocyte-to-macrophage differentiation”. In: **Nature Immunology** 22.6 (2021), pp. 711–722. ISSN: 1529-2908. DOI: 10.1038/s41590-021-00928-y.

[201] Marion Frankenberger, Wilhelm Schwaeble, and Loems Ziegler-Heitbrock. “Expression of M-Ficolin in human monocytes and macrophages”. In: **Molecular Immunology** 45.5 (2008), pp. 1424–1430. ISSN: 0161-5890. DOI: 10.1016/j.molimm.2007.08.014.

[202] Antonella Viola et al. “The Metabolic Signature of Macrophage Responses”. In: **Frontiers in Immunology** 10 (2019), p. 1462. DOI: 10.3389/fimmu.2019.01462.

[203] Kyle Tretina et al. “Interferon-induced guanylate-binding proteins: Guardians of host defense in health and disease”. In: **The Journal of Experimental Medicine** 216.3 (2019), pp. 482–500. ISSN: 0022-1007. DOI: 10.1084/jem.20182031.

[204] William M Schneider, Meike Dittmann Chevillotte, and Charles M Rice. "Interferon-stimulated genes: a complex web of host defenses." English. In: **Annual review of immunology** 32.1 (Mar. 2014), pp. 513–45. ISSN: 0732-0582. DOI: 10.1146/annurev-immunol-032713-120231.

[205] Alberto Mantovani et al. "The chemokine system in diverse forms of macrophage activation and polarization". In: **Trends in Immunology** 25.12 (2004), pp. 677–686. ISSN: 1471-4906. DOI: 10.1016/j.it.2004.09.015.

[206] Ashish Bhattacharjee et al. "IL-4 and IL-13 employ discrete signaling pathways for target gene expression in alternatively activated monocytes/macrophages". In: **Free Radical Biology and Medicine** 54 (2013), pp. 1–16. ISSN: 0891-5849. DOI: 10.1016/j.freeradbiomed.2012.10.553.

[207] Olivier Sordet et al. "Specific involvement of caspases in the differentiation of monocytes into macrophages". In: **Blood** 100.13 (2002), pp. 4446–4453. ISSN: 0006-4971. DOI: 10.1182/blood-2002-06-1778.

[208] C. Teh et al. "M-ficolin is expressed on monocytes and is a lectin binding to N-acetyl-d-glucosamine and mediates monocyte adhesion and phagocytosis of Escherichia coli". In: **Immunology** 101.2 (2000), pp. 225–232. ISSN: 1365-2567. DOI: 10.1046/j.1365-2567.2000.00099.x.

[209] Roman Gunthner and Hans-Joachim Anders. "Interferon-Regulatory Factors Determine Macrophage Phenotype Polarization". In: **Mediators of Inflammation** 2013 (2013), p. 731023. ISSN: 0962-9351. DOI: 10.1155/2013/731023.

[210] Toby Lawrence and Gioacchino Natoli. "Transcriptional regulation of macrophage polarization: enabling diversity with identity". In: **Nature Reviews Immunology** 11.11 (2011), pp. 750–761. ISSN: 1474-1733. DOI: 10.1038/nri3088.

[211] Derin Tugal, Xudong Liao, and Mukesh Jain. "Transcriptional Control of Macrophage Polarization". In: **Arteriosclerosis, Thrombosis, and Vascular Biology** 33.6 (2013), pp. 1135–1144. ISSN: 1079-5642. DOI: 10.1161/atvbaha.113.301453.

[212] Attila Szanto et al. "STAT6 Transcription Factor Is a Facilitator of the Nuclear Receptor PPARg-Regulated Gene Expression in Macrophages and Dendritic Cells". In: **Immunity** 33.5 (2010), pp. 699–712. ISSN: 1074-7613. DOI: 10.1016/j.immuni.2010.11.009.

[213] Jia Xue et al. "Transcriptome-Based Network Analysis Reveals a Spectrum Model of Human Macrophage Activation". In: **Immunity** 40.2 (2014), pp. 274–288. ISSN: 1074-7613. DOI: 10.1016/j.immuni.2014.01.006.

[214] Chynna M. Hendricks et al. "The Interplay of HIV-1 and Macrophages in Viral Persistence". In: **Frontiers in Microbiology** 12 (2021), p. 646447. ISSN: 1664-302X. DOI: 10.3389/fmicb.2021.646447.

[215] J Kimpton and M Emerman. "Detection of replication-competent and pseudotyped human immunodeficiency virus with a sensitive cell line on the basis of activation of an integrated beta-galactosidase gene". In: **Journal of Virology** 66.4 (1992), pp. 2232–2239. ISSN: 0022-538X. DOI: 10.1128/jvi.66.4.2232-2239.1992.

[216] Y C Chen, S Y Wang, and C C King. "Bacterial lipopolysaccharide inhibits dengue virus infection of primary human monocytes/macrophages by blockade of virus entry via a CD14-dependent mechanism." In: **Journal of virology** 73.4 (1999), pp. 2650–7. ISSN: 0022-538X.

[217] Ling-Jun Ho et al. "Dengue Virus Type 2 Antagonizes IFN-a but Not IFN-g Antiviral Effect via Down-Regulating Tyk2-STAT Signaling in the Human Dendritic Cell". In: **The Journal of Immunology** 174.12 (2005), pp. 8163–8172. ISSN: 0022-1767. DOI: 10.4049/jimmunol.174.12.8163.

[218] T R Prestwood et al. “Gamma Interferon (IFN-) Receptor Restricts Systemic Dengue Virus Replication and Prevents Paralysis in IFN- / Receptor-Deficient Mice”. In: **Journal of Virology** 86.23 (2012), pp. 12561–12570. ISSN: 0022-538X. DOI: 10.1128/jvi.06743-11.

[219] Veera Westenius et al. “Highly Pathogenic H5N1 Influenza A Virus Spreads Efficiently in Human Primary Monocyte-Derived Macrophages and Dendritic Cells”. In: **Frontiers in Immunology** 9 (2018), p. 1664. ISSN: 1664-3224. DOI: 10.3389/fimmu.2018.01664.

[220] Igor Kogut et al. “High-efficiency RNA-based reprogramming of human primary fibroblasts”. In: **Nature Communications** 9.1 (2018), p. 745. DOI: 10.1038/s41467-018-03190-3.

[221] Jamie Toombs et al. “Generation of twenty four induced pluripotent stem cell lines from twenty four members of the Lothian Birth Cohort 1936”. In: **Stem Cell Research** 46 (2020), p. 101851. ISSN: 1873-5061. DOI: 10.1016/j.scr.2020.101851.

[222] Kezhen Wang et al. “Interferon-stimulated TRIM69 interrupts dengue virus replication by ubiquitinating viral nonstructural protein 3”. In: **PLoS Pathogens** 14.8 (2018), e1007287. ISSN: 1553-7366. DOI: 10.1371/journal.ppat.1007287.

[223] Alex C Stabell et al. “Dengue viruses cleave STING in humans but not in nonhuman primates, their presumed natural reservoir”. English. In: **eLife** 7 (Mar. 2018), e31919. DOI: 10.7554/elife.31919.

[224] Xun Zhu et al. “IFITM3-containing exosome as a novel mediator for anti-viral response in dengue virus infection”. In: **Cellular Microbiology** 17.1 (2015), pp. 105–118. ISSN: 1462-5814. DOI: 10.1111/cmi.12339.

[225] Nicholas R Meyerson et al. “Nuclear TRIM25 Specifically Targets Influenza Virus Ribonucleoproteins to Block the Onset of RNA Chain Elongation”. English. In: **Cell**

Host & Microbe 22.5 (Nov. 2017), 627–638.e7. ISSN: 1931-3128. DOI: 10.1016/j.chom.2017.10.003.

[226] Tien-Ying Hsiang, Chen Zhao, and Robert M Krug. "Interferon-Induced ISG15 Conjugation Inhibits Influenza A Virus Gene Expression and Replication in Human Cells". In: **Journal of Virology** 83.12 (2009), pp. 5971–5977. ISSN: 0022-538X. DOI: 10.1128/jvi.01667-08.

[227] Xiuyan Wang, Ella R Hinson, and Peter Cresswell. "The Interferon-Inducible Protein Viperin Inhibits Influenza Virus Release by Perturbing Lipid Rafts". English. In: **Cell Host & Microbe** 2.2 (Aug. 2007), pp. 96–105. ISSN: 1931-3128. DOI: 10.1016/j.chom.2007.06.009.

[228] Marc Beyer et al. "High-Resolution Transcriptome of Human Macrophages". In: **PLoS ONE** 7.9 (2012), e45466. DOI: 10.1371/journal.pone.0045466.

[229] Julie Eggenberger et al. "Type I interferon response impairs differentiation potential of pluripotent stem cells". In: **Proceedings of the National Academy of Sciences of the United States of America** 116.4 (2019), pp. 1384–1393. ISSN: 0027-8424. DOI: 10.1073/pnas.1812449116.

[230] Chun Shik Park et al. "Role of the reprogramming factor KLF4 in blood formation". In: **Journal of Leukocyte Biology** 99.5 (2016), pp. 673–685. ISSN: 0741-5400. DOI: 10.1189/jlb.1ru1215-539r.

[231] Mark W Feinberg et al. "The Kruppel-like factor KLF4 is a critical regulator of monocyte differentiation". In: **The EMBO Journal** 26.18 (2007), pp. 4138–4148. ISSN: 1460-2075. DOI: 10.1038/sj.emboj.7601824.

[232] Mark W. Feinberg et al. "Kruppel-like Factor 4 Is a Mediator of Proinflammatory Signaling in Macrophages*". In: **Journal of Biological Chemistry** 280.46 (2005), pp. 38247–38258. ISSN: 0021-9258. DOI: 10.1074/jbc.m509378200.

[233] Xudong Liao et al. “Krüppel-like factor 4 regulates macrophage polarization”. In: **Journal of Clinical Investigation** 121.7 (2011), pp. 2736–2749. ISSN: 0021-9738. DOI: 10.1172/jci45444.

[234] M. A. A. Kusters et al. “Intrinsic defect of the immune system in children with Down syndrome: a review”. In: **Clinical & Experimental Immunology** 156.2 (2009), pp. 189–193. ISSN: 1365-2249. DOI: 10.1111/j.1365-2249.2009.03890.x.

[235] Glenn A. MacLean et al. “Altered hematopoiesis in trisomy 21 as revealed through in vitro differentiation of isogenic human pluripotent cells”. In: **Proceedings of the National Academy of Sciences** 109.43 (2012), pp. 17567–17572. ISSN: 0027-8424. DOI: 10.1073/pnas.1215468109.

[236] Anouchka P. Laurent, Rishi S. Kotecha, and Sébastien Malinge. “Gain of chromosome 21 in hematological malignancies: lessons from studying leukemia in children with Down syndrome”. In: **Leukemia** 34.8 (2020), pp. 1984–1999. ISSN: 0887-6924. DOI: 10.1038/s41375-020-0854-5.

[237] Stewart R Himes et al. “The Runx1 transcription factor controls CSF-1-dependent and -independent growth and survival of macrophages”. In: **Oncogene** 24.34 (2005), pp. 5278–5286. ISSN: 0950-9232. DOI: 10.1038/sj.onc.1208657.

[238] Joseph G. Azofeifa et al. “Enhancer RNA profiling predicts transcription factor activity”. In: **Genome Research** 28.3 (2018), pp. 334–344. ISSN: 1088-9051. DOI: 10.1101/gr.225755.117.

[239] The Geuvadis Consortium et al. “Transcriptome and genome sequencing uncovers functional variation in humans”. In: **Nature** 501.7468 (2013), pp. 506–511. ISSN: 0028-0836. DOI: 10.1038/nature12531.