UCSF
UC San Francisco Electronic Theses and Dissertations

Title

Genomic, Immunologic, Transcriptional and Functional Interrogation of SARS-CoV-2 and Respiratory Syncytial Virus

Permalink

https://escholarship.org/uc/item/95b0687g

Author

Sunshine, Sara

Publication Date

2022

Peer reviewed|Thesis/dissertation

Genomic, Immunologic, Transcriptional and Functional Interrogation of
SARS-CoV-2 and Respiratory Syncytial Virus

by
Sara Sunshine

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
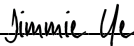DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION
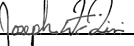of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

_____
Jimmie Ye
Chair

_____
Joseph DeRisi

_____
Melanie Ott

_____

_____
Committee Members

have asked for better teammates and have learned so much from you both. Thank you for your support and comradery. Furthermore, my research would not have been possible without the support of Brittany Worth and Manny De Vera. Thank you Brittany and Manny for the many things you do – all of our research goes smoothly because of you both!

I have also had the great pleasure of working with talented scientists at the Biohub and I thank them all for their feedback and guidance over the years. Dr. Andreas Puschnik, I thank you for your scientific guidance and mentorship throughout my graduate career. Additionally, I extend my gratitude to Dr. Amy Kistler, Dr. Manu Leonetti, Dr. Emily Crawford, Dr. Vida Ahyong, Dr Angela Pisco and Dr. Sandra Schmid.

I would like to thank my dear family and friends for all of their support throughout my graduate studies. I extend my deepest gratitude to my parents. I could not be here without your consistent love and support. Thank you, Nancy and Jeffrey Sunshine, for providing me with everything I've ever needed to succeed, and more.

Lastly, to the best hiking, biking and bakery hopping partner anyone could ever ask for, I thank you from the bottom of my heart. Connor D'Andria, your love and support picked me up during every challenge. I could imagine that living with a virologist during a global pandemic is a bit hectic, thank you for being my rock through it all.

Contributions

Chapter 2

Includes content previously published in:

Pilarowski G, Lebel P, Sunshine S, Liu J, Crawford E, Marquez C, Rubio L, Chamie G, Martinez J, Peng J, Black D, Wu W, Pak J, Laurie MT, Jones D, Miller S, Jacobo J, Rojas S, Rojas S, Nakamura R, Tulier-Laiwa V, Petersen M, Havlir DV, DeRisi J. Performance Characteristics of a Rapid Severe Acute Respiratory Syndrome Coronavirus 2 Antigen Detection Assay at a Public Plaza Testing Site in San Francisco. *J Infect Dis*. 2021 Apr 8;223(7):1139-1144. doi: 10.1093/infdis/jiaa802. PMID: 33394052; PMCID: PMC7799021.

Chapter 3

Includes content previously published in:

Peng J*, Liu J*, Mann SA*, Mitchell AM*, Laurie MT*, Sunshine S*, Pilarowski G*, Ayscue P*, Kistler A, Vanaerschot M, Li LM, McGeever A, Chow ED, Marquez C, Nakamura R, Rubio L, Chamie G, Jones D, Jacobo J, Rojas S, Rojas S, Tulier-Laiwa V, Black D, Martinez J, Naso J, Schwab J, Petersen M, Havlir D, DeRisi J; IDseq Team. Estimation of Secondary Household Attack Rates for Emergent Spike L452R Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Variants Detected by Genomic Surveillance at a Community-Based Testing Site in San Francisco. *Clin Infect Dis*. 2022 Jan 7;74(1):32-39. doi: 10.1093/cid/ciab283. PMID: 33788923; PMCID: PMC8083548.

Chapter 4

Includes content previously published in:

Liu J, Laurie MT, Rubio L, Vazquez SE, Sunshine S, Mitchell AM, Hapte-Selassie M, Mann SA, Pilarowski G, Black D, Marquez C, Rojas S, Lionakis MS, Petersen M, Whitman JD, Jain V,

Anderson M, Havlir D, DeRisi J. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Transmission Dynamics and Immune Responses in a Household of Vaccinated Persons. *Clin Infect Dis*. 2022 Aug 24;75(1):e303-e306. doi: 10.1093/cid/ciac029. PMID: 35037050; PMCID: PMC8807302.

Chapter 5

Includes content previously published in:

Laurie MT, Liu J, Sunshine S, Peng J, Black D, Mitchell AM, Mann SA, Pilarowski G, Zorn KC, Rubio L, Bravo S, Marquez C, Sabatino JJ, Mittl K, Petersen M, Havlir D, DeRisi J. SARS-CoV-2 Variant Exposures Elicit Antibody Responses With Differential Cross-Neutralization of Established and Emerging Strains Including Delta and Omicron. J Infect Dis. 2022 Jun 1;225(11):1909-1914. doi: 10.1093/infdis/jiab635. PMID: 34979030; PMCID: PMC8755395.

Chapter 6

Includes content previously from bioRxiv preprint:

Sunshine S, Puschnik AS, Replogle JM, Laurie MT, Liu J, Zha BZ, Nuñez JK, Byrum JR, McMorrow AH, Frieman MB, Winkler J, Qiu X, Rosenberg OS, Leonetti MD, Ye CJ, Weissman JW, DeRisi JL, Hein MY. Systematic functional interrogation of SARS-CoV-2 host factors using Perturb-seq. bioRxiv **2022**; :2022.07.15.500120.

Chapter 7

Includes contributions by:

Dr. Andreas Puschnik, Dr. Hanna Retallack, Dr. Matthew Laurie, Dr. Jamin Liu, Dr. Duo Peng, Dr. Kristeene Knopp, Dr. Matthew Zinter, Dr. Chun Jimmie Ye and Dr. Joseph DeRisi.

Genomic, Immunologic, Transcriptional and Functional Interrogation of

SARS-CoV-2 and Respiratory Syncytial Virus

Sara Sunshine

Abstract

SARS-CoV-2 and RSV are ubiquitous respiratory viruses, and a leading cause of morbidity and mortality globally. Despite great advances in our biological understanding of these pathogens, there remains a need for multidisciplinary research that improves diagnostics, utilizes genomic surveillance for viral variant detection, characterizes the innate and adaptive immune response to viral infections, and identifies host determinants of infection. In this dissertation, I describe the work that I have contributed to for the systematic interrogation of these aspects of SARS-CoV-2 and RSV biology. Chapters 2 through 5 describe the collaborative studies performed to evaluate a SARS-CoV-2 rapid antigen test (Chapter 2) and to investigate the humoral immune response to SARS-CoV-2 infection (Chapter 3-5). We additionally used Perturb-seq, a single-cell CRISPRi screening method, to functionally interrogate how host genetic perturbations affect SARS-CoV-2 infection dynamics (Chapter 6). The final chapter of this dissertation focuses on RSV. We characterized the transcriptional signatures of RSV infected and bystander activated cells using single-cell RNA sequencing and identified host determinants of RSV infection using a genome-wide CRISPR screening approach. Together, these chapters provide insight into multiple aspects of SARS-CoV-2 and RSV biology, and lay the groundwork for future studies.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

COVID-19 - Coronavirus disease 2019

CRISPR – Clustered regularly interspaced short palindromic repeats

MOI – Multiplicity of infection

NT_{50} – 50% neutralization titer

NT_{90} – 90% neutralization titer

qRT-PCR – Quantitative reverse transcriptase polymerase chain reaction

RSV – Respiratory Syncytial Virus

SARS-CoV-2 – Severe Acute Respiratory Syndrome 2

UMAP – Uniform manifold approximation and projection

Chapter 1: Introduction

Respiratory infections are a global health burden and a leading cause of mortality world-wide [1]. Despite successful vaccine development and distribution campaigns for viral respiratory pathogens such as influenza and COVID-19, there remains a paucity of efficacious and specific antiviral agents for respiratory viruses. Furthermore, as both infection and the subsequent host immune response are thought to damage the host, it remains imperative to understand the innate and adaptive immune response to respiratory viral infections.

Effective management of viral outbreaks requires multidisciplinary and multifaceted strategies including: 1) the development, validation and wide-spread use of pathogen detection methodologies; 2) up-to-date viral genomic surveillance; 3) a detailed understanding of the host response to infection and; 4) identification of host factors that are co-opted for viral infection. While specific introductions are included with each chapter of this thesis, broadly, the work discussed in this dissertation investigates these different aspects of two major respiratory viral pathogens: Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and Respiratory Syncytial Virus (RSV).

SARS-CoV-2

The causative agent of the coronavirus disease 2019 (COVID-19) pandemic, SARS-CoV-2, has claimed millions of lives globally. Rapid vaccine and therapeutic development have sought to alleviate the global health burden of SARS-CoV-2. In this dissertation, we sought to improve SARS-CoV-2 testing, utilize genomic surveillance for viral variant detection, improve our biological understanding of the host innate and humoral response to infection, and to systematically interrogate host determinants of SARS-CoV-2 infection.

Chapter 2 describes a close collaboration of academic (UCSF, UC Berkeley) and community health (Unidos en Salud) programs to evaluate the Abbot BinaxNOW SARS-CoV-2 rapid antigen test [2]. Along with comparing results from this rapid antigen test to qRT-PCR for patient samples, we titrated cultured SARS-CoV-2 and determined the observable threshold of these rapid antigen tests. As SARS-CoV-2 continued to spread globally, we next shifted our efforts to investigate the presence of SARS-CoV-2 variants. In Chapter 3, we again worked as a multidisciplinary team to perform genomic surveillance. In this study, we identified variants circulating in San Francisco, and subsequently determined the viral load and relative household attack rate of the identified strain [3].

We subsequently coupled the genomic surveillance data with both household metadata and methods to investigate the humoral immune response to SARS-CoV-2 infection and vaccination. It was again due to highly collaborative efforts with academic and community organizations that enabled a deep evaluation of the variant specific humoral response. We used a split luciferase antibody sensing assay to detect anti-Nucleocapsid and anti-Spike antibodies [4], as well as a pseudovirus reporter assay [5] to assess the neutralizing capacity of antibodies elicited from vaccination and infection. In Chapter 4, we characterized the SARS-CoV-2 transmission dynamics and the immune response dynamics in a household of vaccinated individuals [6]. In Chapter 5, we investigated cross-neutralization of SARS-CoV-2 variants using the sera from vaccinated, infected, and vaccinated infected individuals [7].

To complement our prior studies, we next investigated the host innate immune response to SARS-CoV-2 infection and performed a systematic interrogation of host factors that are necessary for viral infection. Using Perturb-seq [8–11], an approach that couples single-cell RNA sequencing

and functional genomics, we characterized the heterogeneity of the transcriptional response to infection and identified host factors necessary for infection and bystander activation [12].

RSV

RSV infects an estimated 64 million people and leads to approximately 160,000 deaths annually [13]. RSV is particularly problematic for infants, immunocompromised populations and the elderly with symptoms ranging from coughing and sneezing to more severe wheezing, fever, bronchiolitis and pneumonia. While Palivizumab, an FDA approved monoclonal antibody, may provide protection prophylactically, there remains no effective RSV-specific therapeutic for infected patients [14]. RSV primarily infects airway epithelial cells from the nasal cavity down to the lower respiratory tract [15]. Since epithelial cells are the first line of defense against RSV infection, the temporal and qualitative response of the innate immune system are key determinants of infection outcome [16]. As the first to be infected, epithelial cells set the immunologic tone by inducing a pro-inflammatory response.

In Chapter 7, we investigate the host innate immune response and host factors necessary for RSV infection. We used single-cell RNA sequencing to investigate the transcriptional response of RSV infected and bystander activated cells. To complement this transcriptional data, we subsequently used whole-genome CRISPR/Cas9 screening to identify host determinants of viral infection. We identified host factors important for RSV infection, and comparative analyses revealed overlapping host dependency factors with other RNA viruses.

In summary, this dissertation contributes to our fundamental knowledge about SARS-CoV-2 and RSV biology, and is the culmination of the collaborative and multidisciplinary work that I have proudly contributed to.

Chapter 1 References

1. The top 10 causes of death. Available at: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. Accessed 12 September 2022.

2. Pilarowski G, Lebel P, Sunshine S, et al. Performance Characteristics of a Rapid Severe Acute Respiratory Syndrome Coronavirus 2 Antigen Detection Assay at a Public Plaza Testing Site in San Francisco. J Infect Dis **2021**; 223:1139–1144.

3. Peng J, Liu J, Mann SA, et al. Estimation of Secondary Household Attack Rates for Emergent Spike L452R Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Variants Detected by Genomic Surveillance at a Community-Based Testing Site in San Francisco. Clin Infect Dis Off Publ Infect Dis Soc Am **2021**; 74:32–39.

4. Elledge SK, Zhou XX, Byrnes JR, et al. Engineering luminescent biosensors for point-of-care SARS-CoV-2 antibody detection. Nat Biotechnol **2021**; 39:928–935.

5. Hoffmann M, Kleine-Weber H, Pöhlmann S. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. Mol Cell **2020**; 78:779-784.e5.

6. Liu J, Laurie MT, Rubio L, et al. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Transmission Dynamics and Immune Responses in a Household of Vaccinated Persons. Clin Infect Dis Off Publ Infect Dis Soc Am **2022**; 75:e303–e306.

7. Laurie MT, Liu J, Sunshine S, et al. SARS-CoV-2 Variant Exposures Elicit Antibody Responses With Differential Cross-Neutralization of Established and Emerging Strains Including Delta and Omicron. J Infect Dis **2022**; 225:1909–1914.

8. Adamson B, Norman TM, Jost M, et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. Cell **2016**; 167:1867-1882.e21.

9. Dixit A, Parnas O, Li B, et al. Perturb-seq: Dissecting molecular circuits with scalable single cell RNA profiling of pooled genetic screens. Cell **2016**; 167:1853-1866.e17.

10. Datlinger P, Rendeiro AF, Schmidl C, et al. Pooled CRISPR screening with single-cell transcriptome read-out. Nat Methods **2017**; 14:297–301.

11. Jaitin DA, Weiner A, Yofe I, et al. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. Cell **2016**; 167:1883-1896.e15.

12. Sunshine S, Puschnik AS, Replogle JM, et al. Systematic functional interrogation of SARS-CoV-2 host factors using Perturb-seq. bioRxiv **2022**; :2022.07.15.500120.

13. Respiratory Syncytial Virus (RSV) | NIH: National Institute of Allergy and Infectious Diseases. Available at: https://www.niaid.nih.gov/diseases-conditions/respiratory-syncytial-virus-rsv. Accessed 12 September 2022.

14. de Fontbrune FS, Robin M, Porcher R, et al. Palivizumab treatment of respiratory syncytial virus infection after allogeneic hematopoietic stem cell transplantation. Clin Infect Dis Off Publ Infect Dis Soc Am **2007**; 45:1019–1024.

15. Collins PL, Fearns R, Graham BS. Respiratory Syncytial Virus: Virology, Reverse Genetics, and Pathogenesis of Disease. Curr Top Microbiol Immunol **2013**; 372:3–38.

16. Whitsett JA, Alenghat T. Respiratory epithelial cells orchestrate pulmonary innate immunity. Nat Immunol **2014**; 16:27–35.

Chapter 2: Evaluation of Abbott BinaxNOW rapid antigen test for SARS-CoV-2 detection

This chapter is a reprint of the following publication:

Pilarowski G, Lebel P, Sunshine S, Liu J, Crawford E, Marquez C, Rubio L, Chamie G, Martinez J, Peng J, Black D, Wu W, Pak J, Laurie MT, Jones D, Miller S, Jacobo J, Rojas S, Rojas S, Nakamura R, Tulier-Laiwa V, Petersen M, Havlir DV, DeRisi J. Performance Characteristics of a Rapid Severe Acute Respiratory Syndrome Coronavirus 2 Antigen Detection Assay at a Public Plaza Testing Site in San Francisco. *J Infect Dis*. 2021 Apr 8;223(7):1139-1144. doi: 10.1093/infdis/jiaa802. PMID: 33394052; PMCID: PMC7799021.

Supplemental files are included with original published work.

Abstract

We evaluated the performance of the Abbott BinaxNOW rapid antigen test for coronavirus disease 2019 (Binax-CoV2) to detect virus among persons, regardless of symptoms, at a public plaza site of ongoing community transmission. Titration with cultured severe acute respiratory syndrome coronavirus 2 yielded a human observable threshold between 1.6×10^4-4.3×10^4 viral RNA copies (cycle threshold [Ct], 30.3–28.8). Among 878 subjects tested, 3% (26 of 878) were positive by reverse-transcription polymerase chain reaction, of whom 15 of 26 had a Ct <30, indicating high viral load; of these, 40% (6 of 15) were asymptomatic. Using this Ct threshold (<30) for Binax-CoV2 evaluation, the sensitivity of Binax-CoV2 was 93.3% (95% confidence interval, 68.1%–99.8%) (14 of 15) and the specificity was 99.9% (99.4%–99.9%) (855 of 856).

Main Text

The global pandemic of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection has spread at an unprecedented pace [1] fueled by efficient transmission of infection by the respiratory route, including by asymptomatic and presymptomatic persons. Instances of successful control make use of masking, social distancing, and aggressive testing, tracing, and quarantine [2].

To date, the cornerstone of testing has been reverse-transcription polymerase chain reaction (RT-PCR) examination of respiratory secretions, which has excellent sensitivity and specificity but is expensive and requires sophisticated equipment and highly trained personnel [3]. In practice, these features have often generated testing delays compromising their utility [4]. As a result, there is interest in rapid and economical assays that circumvent these limitations [5]. However, methods that do not include an amplification step are inherently less sensitive; their proper deployment will

7

therefore require a rigorous evaluation of performance characteristics in different epidemiologic settings.

Lateral flow antigen detection diagnostics have been deployed for a variety of infectious diseases including malaria, RSV, and influenza. The Abbott BinaxNOW COVID-19 Ag Card (hereafter referred to as Binax-CoV2) is one such assay that detects viral nucleocapsid (N) protein directly from nasal swab samples. The test requires no instrumentation; results are scored visually and returned within 15 minutes. In August 2020, the Food and Drug Administration issued an emergency use authorization for the diagnosis of SARS-CoV-2 infection in symptomatic patients within 7 days of symptom onset [6]. The US Department of Health and Human Services has distributed 150 million test kits. Given the value of a rapid assessment of infectiousness, there is anticipated use in a broad range of subjects, including those who are asymptomatic. Here we present a systematic examination of the performance characteristics of the Binax-CoV2 test in a community screening setting where testing was offered for symptomatic and asymptomatic subjects.

Methods

Study Population and Specimen Collection

Over 3 days in September 2020, we offered testing in the Mission District, a Latinx-predominant neighborhood, known from prior surveys to have an elevated prevalence of SARS-CoV-2 infection [7, 8]. Walk-up, free testing was conducted at a plaza located at an intersection of the Bay Area-wide subway system (BART) and the San Francisco city bus/streetcar system (MUNI). On the day of testing, participants self-reported symptoms and date of onset, demographics, and contact information, as required by state and federal reporting guidelines. A

laboratory technician performed sequential anterior swab (both nares) for the Binax-CoV2 assay followed by a second swab (both nares) for RT-PCR. Participants were notified of RT-PCR test results. For this study, Binax-CoV2 results were not reported back to study subjects.

Laboratory Testing for SARS-CoV-2

RT-PCR detection of SARS-CoV-2 was performed at the Clinical Laboratory Improvement Amendments–certified laboratory operated by the University of California, San Francisco (UCSF), and the Chan Zuckerberg Biohub, as described elsewhere [9, 10].

Field Testing Using Binax-CoV2 Assay

The Binax-CoV2 assay was performed by technicians on site as described by the manufacturer using the supplied swabs. Each assay was read by 2 independent observers, and a site supervisor served as a tiebreaker. Beginning on day 2 of the study, each Binax-CoV2 assay card was scanned onsite using a color document scanner (CanoScan LIDE 400; Canon). Sample bands were retrospectively quantified from image data. Sample and background regions were localized by offset from the control band, and relative mean pixel intensity decreases were calculated from blue and green channels averaged with respect to background.

Titration of in vitro Cultured SARS-CoV-2 on Binax-CoV2 Cards

SARS-CoV-2 from a UCSF clinical specimen was isolated, propagated and plaqued on Huh7.5.1 cells overexpressing angiotensin-converting enzyme 2 and transmembrane serine protease 2 (TMPRSS2) [11]. Viral titers were determined using standard plaque assays [12]. For titration experiments, SARS-CoV-2 was diluted in Dulbecco phosphate-buffered saline, and 40

µL of each dilution was absorbed onto the supplied swab samples. Images of Binax-CoV2 cards were taken with an Apple iPhone 6. All experiments using cultured SARS-CoV-2 were conducted in a biosafety level 3 laboratory.

N Protein Titration Assay

SARS-CoV-2 N protein (1–419) was expressed in BL21(DE3) Escherichia coli and purified by nickel–nitrilotriacetic acid chromatography, incorporating a 1-mol/L sodium chloride, 50-mmol/L imidazole wash to remove bound RNA. Six concentrations of N protein were tested on 10 lots of Binax-CoV2 kits, and 40 µL of N protein was absorbed onto the provided swab sample.

Ethics Statement

The UCSF Committee on Human Research determined that the study met criteria for public health surveillance. All participants provided informed consent for dual testing.

Results

Binax-CoV2 Performance Using a Titration of in vitro Cultured SARS-CoV-2

To explore the relationship of RT-PCR cycle threshold (Ct), viral load, and the corresponding visual Binax-CoV2 result, a dilution series of laboratory-cultured SARS-CoV-2 with known titers was assayed with both RT-PCR and Binax-CoV2 (Figure 2.1). For this stock of virus, the threshold for detectability by human eye on the Binax-CoV2 assay was between 1.6 and 4.3×10^4 viral copies (100–250 plaque-forming units), corresponding to t values (average of N and E genes) of 30.3 and 28.8, respectively, in this assay.

Community RT-PCR Testing Results

Of the 878 subjects tested, 54% were male, 77% were 18–50 years of age, 81% self-identified as Latinx, and 84% reported no symptoms in the 14 days before testing. Twenty-six persons (3%) were RT-PCR positive; of these, 15 (58%) had a Ct <30, and 6 of the 15 (40%) were asymptomatic. Among asymptomatic individuals with a Ct <30, 4 of 6 developed symptoms within 2 days after testing. Of the 11 persons RT-PCR–positive with a Ct >30, 4 reported symptom onset ≥7 days before testing, 1 reported symptom onset 3 days before testing, and the remainder reported no symptoms.

Comparison of RT-PCR and Binax-CoV2 Testing Results from Community Testing Because the readout of the Binax-CoV2 assay is by visual inspection, results may be subjective, especially when bands are faint or partial. The manufacturer's instructions suggest scoring any visible band as positive. On day 1 of testing, these reading instructions were used and 217 samples tested, of which 214 yielded valid Binax-CoV2 results: 7 of 214 (3.3%) were RT-PCR positive; using the manufacturer's proposed criteria, 5 of these 7 were Binax-CoV2 positive. Of 214, a total of 207 were RT-PCR negative, 9 (4.3%) of which were Binax-CoV2 positive. Thus, using the manufacturer's criteria, 9 of 14 Binax-CoV2–positive tests (64%) in this population of 217 tests had false-positive results (Binax-CoV2 positive/RT-PCR negative). We thought that these initial criteria used on day 1 of testing were insufficient for classifying faint Binax-CoV2 assay bands, resulting in excessive false-positive calls.

On subsequent testing days, we evaluated additional criteria for classifying a band as positive, in consultation with experts from the manufacturer's research staff. Optimal performance occurred when the bands were scored as positive, if they extended across the full width of the strip,

irrespective of the intensity of the band. Updated scoring criteria were implemented by the third day of testing, when a total of 292 tests were administered. Of this total, 283 were RT-PCR negative, all of which scored Binax-CoV2 negative, demonstrating these updated reading criteria markedly alleviated false-positive readings. Of the 292 total day 3 tests, 9 were RT-PCR positive, of which 5 were Binax-CoV2 positive for antigen with these updated scoring criteria. Of the 9 RT-PCR–positive samples, the 4 that were Binax-CoV2 negative had a Ct >30, consistent with our laboratory-observed limit of detection for Binax-CoV2. We find that scoring a test as positive if bands extend across the full width of the strip, irrespective of band intensity, is the least subjective and easiest method to implement in the field, and we have developed a training tool (https://unitedinhealth.org/binax-training).

The results of the 26 RT-PCR–positive individuals identified throughout the 3-day study were stratified by RT-PCR test Ct value and categorized according to Binax-CoV2 result (Figure 2.2). The rapid antigen detection test performed well in samples with higher viral loads: 15 of 16 samples with a Ct<32 were Binax-CoV2 positive (Figure 2.2A). By contrast, none of the 10 samples with a Ct ≥34 were positive by Binax-CoV2 antigen detection. Retrospective image quantification of Binax-CoV2 sample band intensity is correlated with RT-PCR Ct values for those individuals (Figure 2.2B). In each case, the corresponding image is shown to demonstrate the correspondence between RT-PCR and the visual result (Figure 2.2C).

Sensitivity and Specificity

RT-PCR is considered a reference standard [3] and, in the RT-PCR assay used in this study, has a limit of detection of 100 viral RNA copies/mL. Direct antigen assays are inherently not as sensitive as RT-PCR. In the context of community based testing, we defined a threshold for high

virus levels corresponding to the range of highest probability of transmissibility: a Ct of 30, which corresponds to a viral RNA copy number of approximately 1.9×104 in this assay [10, 13]. Using this Ct <30 case definition and 95% confidence intervals (CIs), the sensitivity of the Binax-CoV2 was 93.3% (95% CI, 68.1%–99.8%) (14 of 15), and the specificity was 99.9% (99.4%–99.9%) (855 of 856). Adjusting the threshold to a more conservative Ct value of 33 (2.6×103 viral RNA copies), the sensitivity was 93.8% (95% CI, 69.8%–99.8%) (15 of 16), and the specificity was 100% (99.6%–100%) (855 of 855). Without a Ct threshold, the sensitivity of the Binax-CoV2 assay was (57.7%; 95% CI, 36.9%–76.6%) (15 of 26), and the specificity was (100%; 99.6%–100%) (845 of 845). Given that the Binax-CoV2 assay detects infected individuals with high levels of virus (>104), the sensitivity of the assay in the absence of a threshold will largely depend on the viral kinetics within the testing population. Sensitivity and specificity calculations were completed with the final scoring criteria, using retroactive Binax-CoV2 scores from images covering all 3 study days.

Evaluation of Binax-CoV2 Lot-to-Lot Variation

We quantified lot-to-lot variability in 10 different lots of Binax-CoV2 card tests using a dilution series of N protein. (Figure 2.3). At protein concentrations of ≥17.2 ng/mL, a sample band was detected in all lots and thus would not affect the outcome of this binary assay (Figure 2.3A).

Discussion

The data reported here describe the performance characteristics of the Binax-CoV2 antigen detection kit in the context of community testing including asymptomatic subjects. These results indicate a clear relationship between relative viral load and test positivity and provide a practical,

real-world criterion to assist calling results in this setting. We found that small training modifications reduced the presence of false-positives, a legitimate concern for the rollout of these tests.

The currently approved emergency use authorization for the Binax-CoV2 assay specifies use only in symptomatic individuals. The results presented here suggest that the Binax-CoV2 test should not be limited to symptomatic testing alone. Many asymptomatic individuals have high viral loads (corresponding to low Ct values) and, therefore, have a high probability of being infectious and transmitting the virus, a feature and likely driver of the pandemic that we and others have observed previously [7, 14]. Limiting use of Binax-CoV2 to symptomatic individuals would have missed nearly half of the SARS-CoV-2 infections in the current study.

Furthermore, the impact of testing on forward transmission is hampered by long wait times. Our group reported previously that in the community setting, by the time a person is tested, counseled, and situated under isolation conditions, the effective isolation period is often nearly over [8]. This is particularly true for many communities of color, where reported delays in accessing tests and results are even longer [4, 15]. Rapid tests could reduce these delays and maximize the time of effective isolation. Limitations of our study include its cross-sectional design and the overall small number of RT-PCR positive cases. Additional field performance of this assay is needed and will help inform optimal use strategies. We recommend evaluating the Binax-CoV2 assay side by side with RT-PCR in each context where it will be used before using Binax-CoV2 without RT-PCR.

During the early stages of infection, viral load may be too low to detect by direct antigen assays such as Binax-CoV2. This inherent lower sensitivity may be offset by faster turn-around and higher frequency of testing, with overall lower cost, relative to RT-PCR methods. That said,

for persons who present with a high index of suspicion of coronavirus disease 2019 and a negative Binax-CoV2 result, the test should be complemented with RT-PCR or a repeated Binax-CoV2 test at a later time to make sure cases are not missed.

In summary, under field conditions with supplementary technician training, the Binax-CoV2 assay accurately detected SARS-CoV-2 infection with high viral loads in both asymptomatic and symptomatic individuals. The Binax-CoV2 test could be a valuable asset in an arsenal of testing tools for the mitigation of SARS-CoV-2 spread, as rapid identification of highly infectious individuals is critical.

Supplementary Data

Supplementary materials are available at The Journal of Infectious Diseases online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.
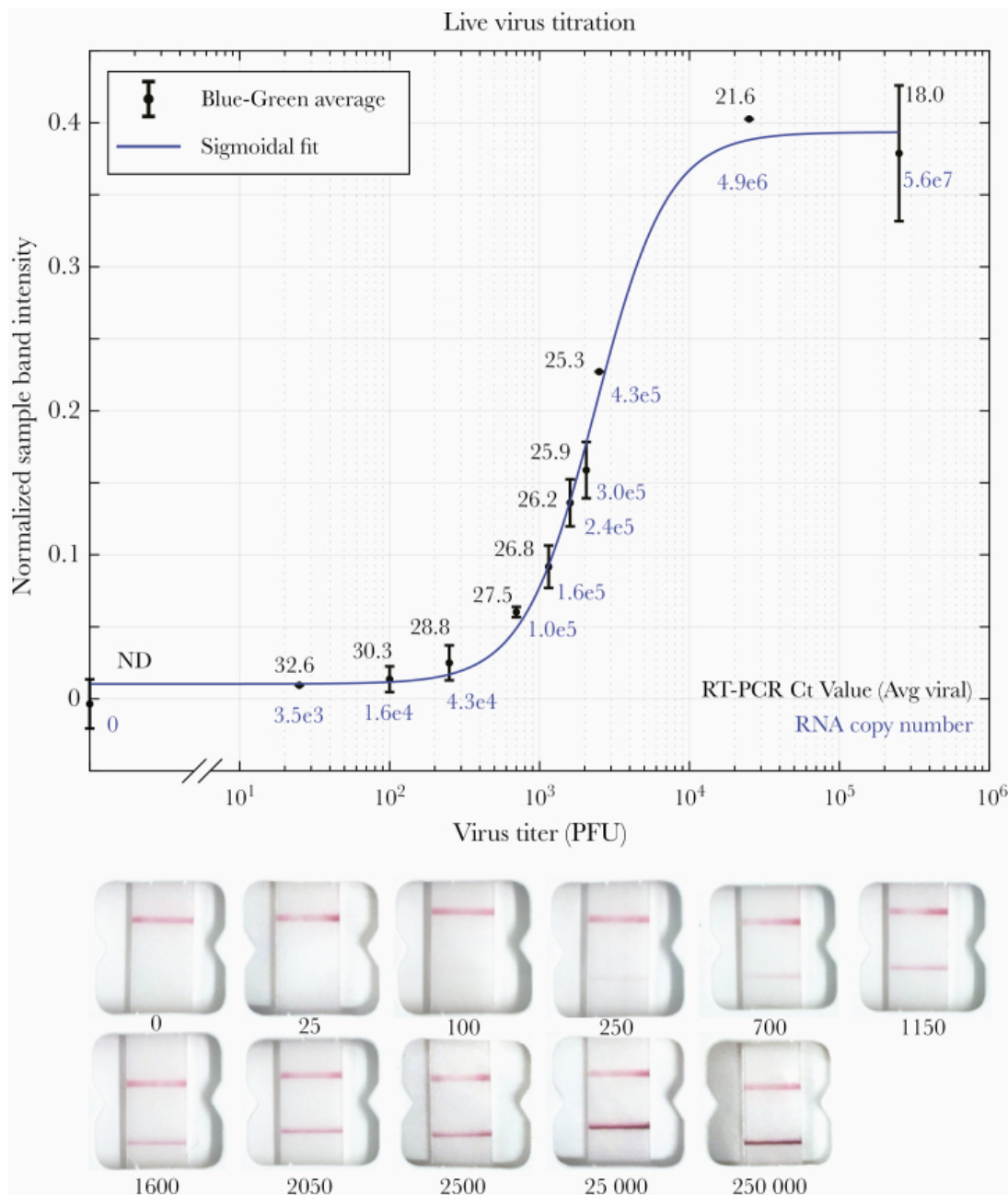
Figure 2.1. Titration of in vitro grown severe acute respiratory syndrome coronavirus 2 and detection with Binax-CoV2 assay. *Top*, Normalized Binax-CoV2 sample band intensity (blue-green average) for cards loaded with a known amount of virus. Error bars represent standard deviation of sample band intensity of technical replicates. Reverse-transcription polymerase chain reaction (RT-PCR) testing was performed at the CLIAHUB consortium [10]. Corresponding RT-PCR cycle threshold (Ct) values (average of N and E gene probes) are shown in black, and the corresponding RNA copy numbers in blue. Note that Ct and genome copy number correlation varies by RT-PCR platform. *Bottom*, Representative card images from each data point. Abbreviation: PFUs, plaque-forming units.
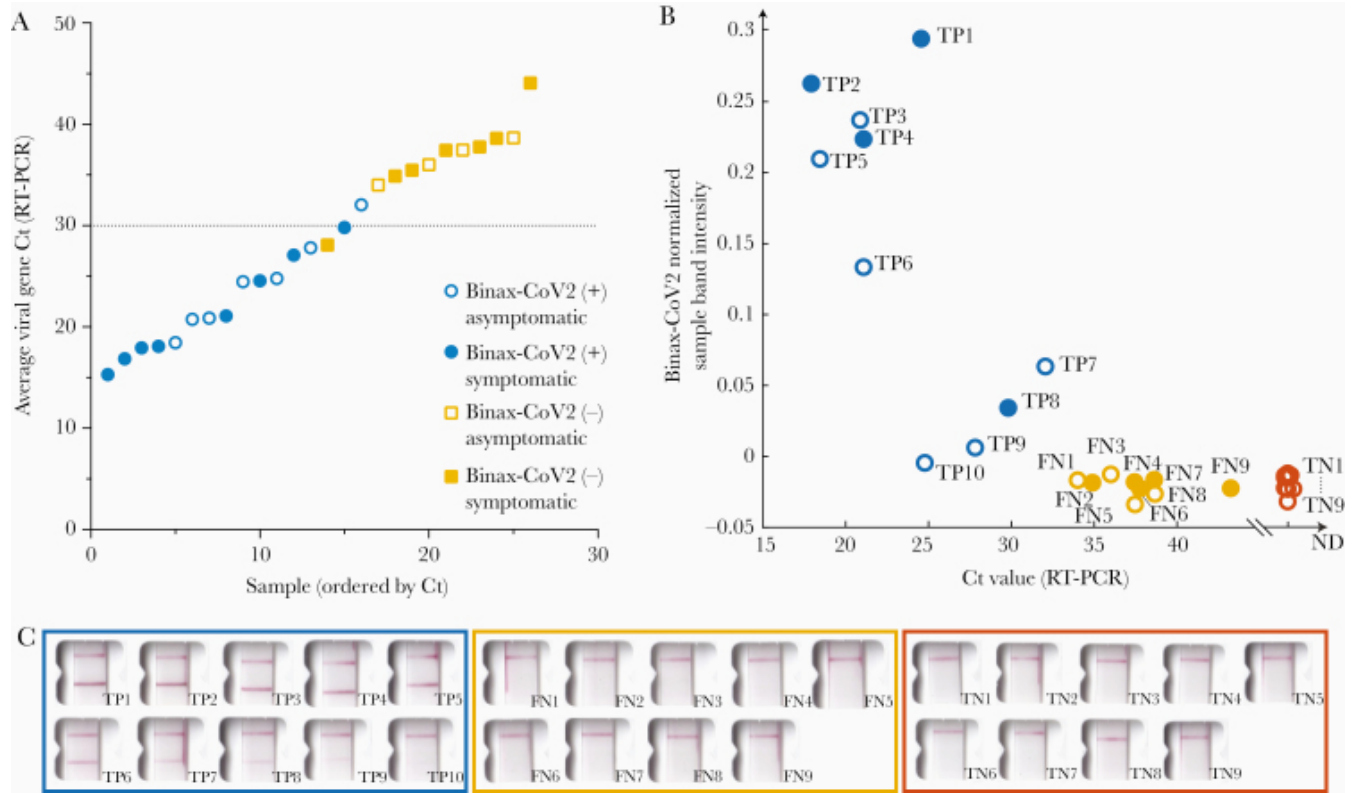
Figure 2.2. Comparison of Binax-CoV2 test with quantitative reverse-transcription polymerase chain reaction (RT-PCR) test. *A*, Average viral cycle threshold (Ct) values from all 26 RT-PCR–positive individuals from the community study, plotted in ascending order. Blue circles indicate Binax-CoV2–positive samples; yellow squares, Binax-CoV2–negative samples. Open symbols represent individuals who were asymptomatic on the day of the test and filled symbols, those who reported symptoms on that day. *B*, Normalized sample band signal from retrospective image analysis of Binax-CoV2 cards was plotted as a function of Ct value for all available scanner images (19 of 26 RT-PCR–positive samples and a random subset of RT-PCR–negative samples). Binax-CoV2 true-positives are shown in blue and labeled TP; false-negatives, shown in yellow and labeled FN; and true-negatives, shown in red and labeled TN. *C*, Corresponding Binax-CoV2 card images from the data in *B*.
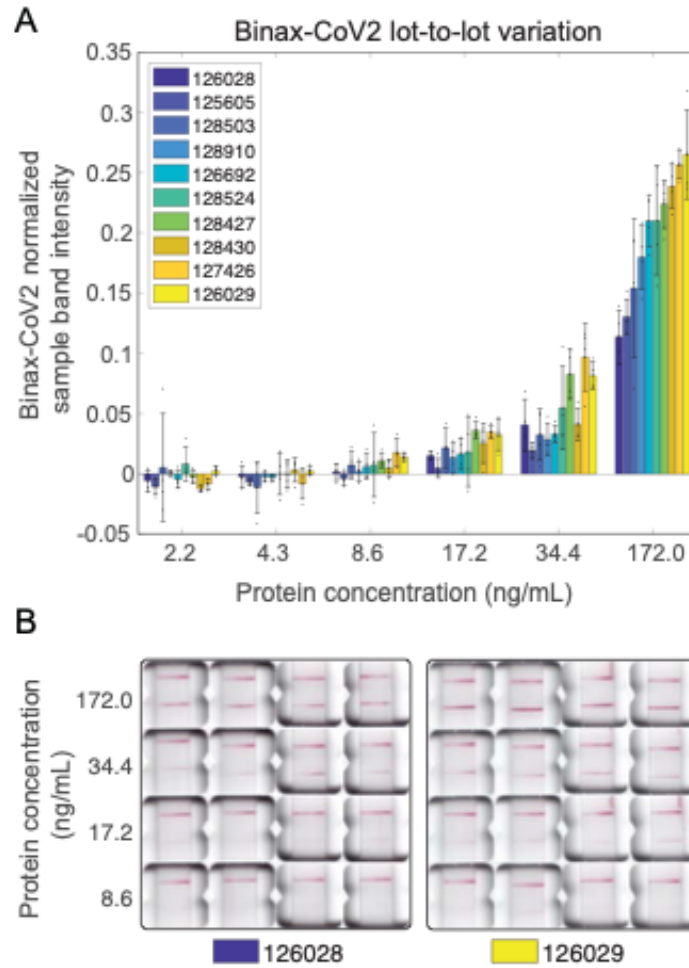
Figure 2.3. Variability of signal intensity in Binax-CoV2 card lots. (A) Normalized sample band signal intensity of Binax-CoV2 cards from different lots run with a dilution series of purified SARS-CoV-2 N protein with known concentration. N=4 cards per lot per concentration. Each point represents one card. (B) Images of each card test for the highest (126029) and lowest (126028) performing lots.

Chapter 2 References

1. Johns Hopkins Coronavirus Resource Center. COVID-19 dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). https://coronavirus.jhu.edu/map.html. Accessed 23 October 2020.

2. Baker MG, Wilson N, Anglemyer A. Successful elimination of Covid-19 transmission in New Zealand. *N Engl J Med* 2020; 383:e56.

3. Esbin MN, Whitney ON, Chong S, Maurer A, Darzacq X, Tjian R. Overcoming the bottleneck to widespread testing: a rapid review of nucleic acid testing approaches for COVID-19 detection. *RNA* 2020; 26:771–83.

4. Chwe H, Quintana A, Lazer D, et al.. The state of the nation: a 50-state COVID-19 survey. Report #17: COVID-19 test result times. October 2020. http://www.kateto.net/covid19/COVID19%20CONSORTIUM%20REPORT%2017%20TESTING%20OCT%202020.pdf. Accessed 23 October 2020.

5. Mina MJ, Parker R, Larremore DB. Rethinking Covid-19 test sensitivity—a strategy for containment. *N Engl J Med* 2020; 383:e120.

6. Abbott Diagnostics. Abbott BinaxNOW COVID-19 Ag package insert, version 1.6. 2020. https://www.fda.gov/media/141570/download. Accessed 20 October 2020.

7. Chamie G, Marquez C, Crawford E, et al.. SARS-CoV-2 community transmission disproportionately affects Latinx population during shelter-in-place in San Francisco [published online ahead of print August 21, 2020]. *Clin Infect Dis* 2020; doi: 10.1093/cid/ciaa1234

8. Kerkhoff AD, Sachdev D, Mizany S, et al.. Evaluation of a novel community-based COVID-19 'Test-to-Care' model for low-income populations. *PLoS One* 2020; 15:e0239400.

9. Crawford ED, Acosta I, Ahyong V, et al.. Rapid deployment of SARS-CoV-2 testing: the CLIAHUB. *PLoS Pathog* 2020; 16:e1008966.

10. Vanaerschot M, Mann SA, Webber JT, et al.. Identification of a polymorphism in the N gene of SARS-CoV-2 that adversely impacts detection by RT-PCR. *J Clin Microbiol* 2020; 59:e02369-20.

11. Wang R, Simoneau CR, Kulsuptrakul J, et al.. Functional genomic screens identify human host factors for SARS-CoV-2 and common cold coronaviruses. *bioRxiv* [Preprint: not peer reviewed]. 24 September 2020. Available from: https://www.biorxiv.org/content/10.1101/2020.09.24.312298v1.

12. Honko AN, Storm N, Bean DJ, Vasquez JH, Downs SN, Griffiths A. Rapid quantification and neutralization assays for novel coronavirus SARS-CoV-2 using Avicel R RC-591 semi-solid overlay. 2020. https://www.preprints.org/manuscript/202005.0264/v1. Accessed 23 October 2020.

13. Wölfel R, Corman VM, Guggemos W, et al.. Virological assessment of hospitalized patients with COVID-2019. *Nature* 2020; 581:465–9.

14. Oran DP, Topol EJ. Prevalence of asymptomatic SARS-CoV-2 infection : a narrative review. *Ann Intern Med* 2020:362–7.

15. Kim HN, Lan KF, Nkyekyer E, et al.. Assessment of disparities in COVID-19 testing and infection across language groups in Seattle, Washington. *JAMA Netw Open* 2020; 3:e2021213.

Chapter 3: Investigation of secondary attack rates of SARS-CoV-2 emerging variants in San Francisco

This chapter is a reprint of the following publication:

*These authors contributed equally to this work

Supplemental files are included with original published work.

Abstract

Background

      Sequencing of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) viral

genome from patient samples is an important epidemiological tool for monitoring and responding

to the pandemic, including the emergence of new mutations in specific communities.

Methods

      SARS-CoV-2 genomic sequences were generated from positive samples collected, along

with epidemiological metadata, at a walk-up, rapid testing site in the Mission District of San

Francisco, California during 22 November to 1 December, 2020, and 10–29 January 2021.

Secondary household attack rates and mean sample viral load were estimated and compared

across observed variants.

Results

      A total of 12 124 tests were performed yielding 1099 positives. From these, 928 high-

quality genomes were generated. Certain viral lineages bearing spike mutations, defined in part

by L452R, S13I, and W152C, comprised 54.4% of the total sequences from January, compared

to 15.7% in November. Household contacts exposed to the "California" or "West Coast" variants

(B.1.427 and B.1.429) were at higher risk of infection compared to household contacts exposed

to lineages lacking these variants (0.36 vs 0.29, risk ratio [RR] = 1.28; 95% confidence interval

[CI]: 1.00–1.64). The reproductive number was estimated to be modestly higher than other

lineages spreading in California during the second half of 2020. Viral loads were similar among

persons infected with West Coast versus non-West Coast strains, as was the proportion of

individuals with symptoms (60.9% vs 64.3%).

Conclusions

The increase in prevalence, relative household attack rates, and reproductive number are consistent with a modest transmissibility increase of the West Coast variants.

Main Text

Genomic surveillance during the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic is a critical source of situational intelligence for epidemiological control measures, including outbreak investigations and detection of emergent variants [1]. Countries with robust, unified public health systems and systematic genomic surveillance have been able to rapidly detect SARS-CoV-2 variants with increased transmission characteristics, and mutations that potentially subvert both naturally acquired or vaccination-based immunity (eg, COVID-19 Genomics UK Consortium). Examples include the rapidly spreading B.1.1.7 lineage documented in the United Kingdom and the B.1.351 lineage described from South Africa, or the P.1/P.2 lineages that harbor the spike E484K mutation that is associated with reduced neutralization in laboratory experiments [2–5].

In the United States, genomic surveillance is sparse relative to the number of confirmed cases (27.8 million as of 20 February 2021), with 123 672 genomes deposited in the GISAID database, representing only 0.4% of the total reported cases. Despite the low rates of US genomic surveillance, independent local programs and efforts have contributed to our understanding of variant emergence and spread [6–8]. The appearance of new nonsynonymous mutations highlight the utility of this approach in the United States [9].

Genomic sequencing of SARS-CoV-2 in California has predominantly been conducted by academic researchers and non-profit biomedical research institutions (eg, the Chan Zuckerberg

Biohub and the Andersen Lab at the Scripps Research Institute) in conjunction with state and local public health partners. These efforts identified an apparent increase in the prevalence of lineages B.1.427 and B.1.429 ("California" or "West Coast" variant), which share S gene nonsynonymous mutations at sites 13, 152, 452, and 614, during December 2020 to February 2021 when California was experiencing the largest peak of cases observed during the pandemic. Although the cluster of mutations was first observed in a sample from May 2020, these variants rose from representing <1% of the consensus genomes recovered from California samples collected in October 2020 (5/546; 0.91%) to over 50% of those collected during January 2021 (2309/4305; 53.6%; GISAID accessed 20 February 2021).

The majority of sequencing efforts in the United States utilize samples from symptomatic individuals or outbreaks, introducing selection bias making interpretation of trends, such as the rise in lineage prevalence, complex. Furthermore, clinical remnant samples are most often delinked from case information, thus eliminating the possibility of evaluating genotypes with detailed household information, and other metadata useful for investigation of transmission dynamics.

Sequencing cases identified during intensive, longitudinal community-based testing may help address both limitations. Here we describe an investigation of the prevalence of the West Coast variants as well as other variants among persons tested at a community testing site situated in the Mission District of San Francisco, a neighborhood with high coronavirus disease 2019 (COVID-19) incidence, during 2 periods: 22 November to 1 December 2020 and 10–29 January 2021. Using metadata collected at the testing site and supplementary household testing, we estimated secondary household attack rate with respect to viral genotype to evaluate relative transmissibility of identified variants.

Methods

Study Setting and Population

Over 22 November to 1 December 2020 and 10–29 January 2021, BinaxNOW[TM] rapid antigen tests were performed at the 24th & Mission BART (public transit) station in the Mission District of San Francisco, a setting of ongoing community transmission, predominantly among Latinx persons [10, 11]. Tests for SARS-CoV-2 were performed free of charge on a walk-up, no-appointment basis, including persons ≥1 year of age and regardless of symptoms, through "Unidos en Salud," an academic, community (Latino Task Force) and city partnership. Certified lab assistants collected 2 bilateral anterior nasal swabs. The first was tested with BinaxNOW[TM], immediately followed by a separate bilateral swab for SARS-CoV-2 genomic sequencing [11, 12]. Results were reported to participants within 2 hours, and all persons in a household (regardless of symptom status) corresponding to a positive BinaxNOW case were offered BinaxNOW testing. All persons testing BinaxNOW positive were offered participation in longitudinal Community Wellness Team support program [13, 14].

SARS-CoV-2 Genomic Sequence Recovery and Consensus Genome Generation

SARS-CoV-2 genomes were recovered using ARTIC Network V3 primers [15] and sequenced on an Illumina NovaSeq platform. Consensus genomes generated from the resulting raw.fastq files using IDseq [16] were used for subsequent analysis. Full details are included in Supplementary materials.

Household Attack Rate Analyses

Households (n = 328) tested in January and meeting the following inclusion criteria were eligible for secondary attack rate analyses: 1) ≥1 adult (aged ≥ 18 years) with a positive BinaxNOW result; 2) ≥1 case in household sequenced; and 3) ≥2 persons tested with BinaxNOW during the study period. Households in which sequences represented both West Coast and non-West Coast variants were excluded (n = 9). The index was defined as the first adult to test positive. Crude household attack rates, stratified by variant classification, were calculated as i) the proportion of positive BinaxNOW results among tested household contacts; and ii) the mean of the household-specific secondary attack rate, with 95% confidence interval (CI) based on cluster-level bootstrap. Generalized estimating equations were used to fit Poisson regressions, with cluster-robust standard errors and an exchangeable working covariance matrix. Because symptoms and disease severity may be affected by strain, these factors were not included in the a priori adjustment set. We evaluated for overdispersion [17] and conducted sensitivity analyses using targeted maximum likelihood estimation (TMLE) combined with Super Learning to relax parametric model assumptions; influence curve-based standard error estimates used household as the unit of independence [18].

Bayesian Phylogenetic Analysis

We compared the growth rates of B.1.427 and B.1.429 Phylogenetic Assignment of Named Global Outbreak (PANGO) lineages against 2 other lineages, B.1.232 and B.1.243, that had been circulating in California during the latter half of 2020. To do this, we built a Bayesian phylogeny for each lineage in BEAST v.1.10.4 and estimated the effective population size over time using the Bayesian SkyGrid model. We fit an exponential model to the median SkyGrid curve and

inferred the reproductive numbers based on the exponential growth rates and generation time estimates from literature. Full analysis details are included in Supplementary materials.

Ethics Statement

The UCSF Committee on Human Research determined that the study met criteria for public health surveillance. All participants provided informed consent for dual testing.

Results

Low-Barrier SARS-CoV-2 Testing and Sequencing

From 22 November to 1 December 2020, 3302 rapid direct antigen tests were performed on 3122 unique individuals; sample characteristics from this testing have been previously described [11]. From 10–29 January, using identical methods, 8822 rapid direct antigen tests were performed on 7696 unique individuals, representing 5239 households; household attack rate analyses were restricted to January samples, described here (Supplementary 3.1).

Test subjects originated from addresses in 8 Bay Area counties, indicating a wide catchment area (Figure 3.1). During this time period, there were 885 (10.0%) samples from 863 unique persons that were BinaxNOW positive for SARS-CoV-2 infection. From this set, a total of 80 samples were sequenced for the S gene only, of which 58 had S gene coverage over 92%. In addition, full SARS-CoV-2 genome sequencing was attempted on a total of 775 samples, of which 737 (95%) samples resulted in a genome coverage over 92% (Supplementary Table 3.2, sequences deposited in GISAID). These 986 samples, together with an additional 191 SARS-CoV-2 genome sequences generated from the same testing site during the period of 22 November to 1 December 2020 [11, 19] had adequate coverage of the full genome or spike protein for further analysis based on S

gene sequence (Supplementary Table 3.3). Classification as either a West Coast variant or a non-West Coast variant was determined for 846 of all samples sequenced.

Similar to previous observations in San Francisco [20], full length sequences were distributed among the major clades (Supplementary Figure 3.1) [21]. Notably, mutations at spike position 501 were not observed, and thus no instances of the B.1.1.7 strain or any other strain bearing the N501Y mutation were detected in any sample during this period in January 2021. A single individual was found to have been infected with the P.2 strain, which carries the spike E484K mutation and was described in Brazil from a reinfection case [5]. This mutation has been associated with decreased neutralization in laboratory experiments [2, 4].

We observed SARS-CoV-2 genome sequences that belonged to PANGO lineages B.1.427 and B.1.429, both of which share a trio of recent mutations in the spike protein (S13I, W152C, and L452R) (Figure 3.2). These lineages are separated by differing mutations ORF1a and ORF1b, including ORF1b:P976L and ORF1a:I4205V, respectively. Sequencing of 191 viral genomes from 22 November to 1 December 2020 revealed that sequences carrying this trio of mutations represented only 15.7% of the total. A trend of increasing frequency was observed on a daily basis during the January testing period (Figure 3.2A), and the frequency of these lineages were observed to have increased to 54.4% of the total, representing an increase of more than 3-fold in approximately 1.5 months (Figure 3.2B, 2C).This increase in frequency is consistent with an expansion of viruses more broadly in California carrying these same mutations [23].

Additional nonsynonymous mutations were observed throughout the genome, including 108 unique non-synonymous mutations in the spike gene, several within functionally-significant regions of the protein (Figure 3.2C, Supplementary Table 3.3). Twelve unique mutations were observed in the receptor binding domain, most of which have yet to be investigated for functional

effects. Additionally, 8 unique mutations were found adjacent to the polybasic furin cleavage site at the S1/S2 junction, which is reported to have a potential role in determination of virulence and host cell tropism [24–27]. Moderately prevalent mutations were observed at spike position 681 (P681H, n = 34 and P681R, n = 1), which is within the furin recognition site, and at spike position 677, where 2 different amino acid substitutions were observed in this cohort (Q677H, n = 22 and Q677P, n = 11). Multiple mutations at both of these sites have been previously observed [9].

Disease Severity

The SARS-CoV-2 reverse transcription polymerase chain reaction (RT-PCR) cycle thresholds (Ct) for nasal swab samples from which whole genomes corresponding to the West Coast variant were recovered were compared to parallel non-West Coast variant samples. Mean Ct values did not differ significantly between persons infected with West Coast (mean Ct 23.56; interquartile range [IQR] 6.4) versus non-West Coast (mean Ct 23.67; IQR 7.8) strains (95% CI: −.77 to .50, *P*-value = .67) (Supplementary Figure 2, Supplementary Table 3.2). The proportion of individuals with symptoms was similar among persons infected with West Coast (273/448, 60.9%) versus non-West Coast (250/389, 64.3%) strains. Among 364 sequenced cases with longitudinal follow-up by the Community Wellness Team, 4 (1.1%) were hospitalized (3/183, and 1/181, for West Coast and non-West Coast, respectively).

Household Secondary Attack Rate

A total of 328 households met inclusion criteria for evaluation of secondary attack rate; of these, 9 households had individuals with mixed strains and thus were excluded from analyses. Among the remaining 319 households, characteristics including race/ethnicity, ages of other

household members, household size, density, and location were similar, regardless of whether the members were positive for West Coast or non-West Coast variants. (3.1, Supplementary Table 4).

The 319 index cases had a total of 1241 nonindex household members; of these, 867 (69.9%) had a BinaxNOW test result available (452/658 [68.7%] for West Coast variant households; 415/583 [71.2%] of non-West Coast variant households). A total of 35.6% (161/452) of household contacts exposed to the West Coast variant tested BinaxNOW positive (33.2%, 78/235 for B1.427; 40.3%, 79/196 for B.1.429), whereas 29.4% (122/415) of contacts exposed to non-West Coast variant tested positive (Table 3.2). Secondary cases were identified a median of 1 day after index cases (IQR 0–4).

Based on unadjusted Poisson regression with cluster-robust standard errors, household contacts exposed to the West Coast variant had an estimated 28% higher risk of secondary infection, compared to household contacts exposed to a non-West Coast variant (RR: 1.28, 95% CI: 1.00–1.64, P-value = .05). When exposure to West Coast variants was disaggregated by B.1.427 and B.1.429, corresponding risks of secondary infections relative to exposure to non-West Coast variants were 1.19 (95% CI: .89–1.59, P-value = .20) and 1.43 (95% CI: 1.07–1.91, P-value = .02), respectively. Dispersion ratios were >0.9 in all regression analyses. Estimated relative risks of infection after household exposure to West Coast versus non-West Coast variants were similar after adjustment for household and individual-level characteristics of secondary contacts (adjusted risk ratio [aRR]: 1.25, 95% CI: .98–1.59, P-value: .07 for West Coast vs non-West Coast variants; aRR: 1.19, 95% CI: .90–1.59, P-value = .20 and aRR: 1.36, 95% CI: 1.01–1.83, P-value = .04 for B.1.427 and B.1.429, respectively). Relative attack rates were generally similar when stratified by household characteristics and by the characteristics of secondary contacts (Table 3.3); secondary attack rates among children aged <12 years were 51.9% (41/79) and 39.7% (31/78) when exposed

to West Coast and non-West Coast strains, respectively. Sensitivity analyses in which parametric assumptions were relaxed using TMLE and Super Learning yielded similar estimates (Supplementary Table 5).

Estimation of Reproductive Number

Using Bayesian phylogenetic analysis, we estimated the reproductive number to be 1.27 (95% CI: 1.10–1.46) for B.1.427 and 1.18 (95% CI: 1.05–1.32) for B.1.429 during the second half of 2020. These values were slightly higher than 2 other lineages spreading in California during the same time period: 1.12 (95% CI: 1.10–1.14) for B.1.232, and 1.02 (95% CI: .98–1.05) for B.1.243. As the reproductive numbers are very similar and were calculated from the median SkyGrid estimates, we cannot conclude any statistically significant differences between the lineages.

Discussion

We monitored SARS-CoV-2 viral variants by genomic sequencing and integration of metadata from households at a community based "test-and-respond" program. We found that the West Coast variants (PANGO lineages B.1.427 and B.1.429) increased in prevalence relative to wild type from November to January in the San Francisco Bay Area among persons tested in the same community-based location. These data extend and confirm prior observations from convenience, outbreak, and clinical samples reporting apparent increases in relative prevalence of the West Coast variants [23].

Household secondary attack rates of the West Coast variants were modestly higher than for non-West Coast variants, suggesting the potential for increased transmissibility. The West Coast variants compromise two closely related lineages (B.1.427 and B.1.429) that share identical

sets of mutations in the spike protein but differ by additional synonymous and nonsynonymous mutations in other genes. Although the frequency of both lineages increased in this study and in California more widely [23], and the estimated increase in risk of secondary household infection relative to non-West Coast variants was fairly consistent across lineages, the point estimate was somewhat higher for B.1.429. Although moderate compared to increased transmissibility of other previously identified variants, even small increases in transmissibility could contribute to a substantial increase in cases, particularly in the context of reproductive numbers just below 1. Although this finding may be due to chance, future work should continue to monitor individual lineages.

The household attack rate observed here was higher than that reported in a recent global meta-analysis [28], even for the non-West Coast variants. It was similar to, or lower than, attack rates reported in other US settings. Prior US reports, however, were based on substantially smaller sample sizes.

Our findings that the West Coast variants increased in relative prevalence and had higher household secondary attack rates potentially suggest higher transmissibility. However, the West Coast variant has been detected in multiple locations and has been detected since May 2020 in California without relative expansion until the peak associated with the holiday season of November–January. Using Bayesian phylogenetic analysis, the estimated reproductive number for both West Coast lineages was found to be modestly higher than other circulating lineages.

We found no significant differences in viral load (using Ct) between West Coast and non-West Coast variants (Supplementary Figure 2), and recorded hospitalizations (n = 5/388) remained rare, despite the West Coast variant representing 54.4% of positive cases. This highlights the

importance of studying walk-up populations, whether they are symptomatic or asymptomatic, as hospitalized populations often are confounded by comorbidities and subject to selection bias.

At the time of this sampling, no instances of B.1.1.7, or independent N501Y mutations were detected in our sample population of 830, despite sporadic observations elsewhere in CA (approximately 3% [69/2423] of genomes reported in California during the January study period; accessed from GISAID 24 February 2021), suggesting that introductions of B.1.1.7 have been rare in this catchment area, despite high SARS-CoV-2 incidence [29]. A single case of the P.2 variant, which carries the E484K mutation [2], was detected in this study. Surprisingly, this case did not have a travel history, highlighting the risk of cryptic transmission.

In addition to the mutations associated with spike L452R in the West Coast variants, we observed, at lower frequencies, other mutations of interest, such as those at spike positions 677 and 681, both of which have been reported previously on their own [9].

This study has several limitations. First, testing was conducted at a walk-up testing site, and thus these are inherently convenience samples; however, this would not be expected to impose a differential selection bias for those with or without any particular variant. Second, clear classification of the index case was not always possible, particularly when multiple adults from a household tested positive on the same date; furthermore, secondary household attack rate calculations do not account for potential external sources of infection other than the index case. However, the relative risk of secondary infection from household exposure to West Coast versus non-West Coast variants was similar among children, a group less likely to have been misclassified as non-index or to be exposed to external infection. Third, household testing coverage was incomplete and, in some cases, consisted of only a single follow-up test; this might contribute to

an underestimate (or overestimate) of secondary attack rate, and although we again have no reason to suspect differential ascertainment by strain, this could bias estimates of relative risk.

The occurrence of variants in SARS-CoV-2 was always expected; however, it is often difficult to understand the clinical and epidemiological importance of any given single or set of co-occurring mutations. Although further epidemiological and laboratory experiments will be required to fully understand the community impact and mechanistic underpinnings of each variant, it is clear that enhanced genomic surveillance paired with community engagement, testing, and response capacity is an important tool in the arsenal against this pandemic.
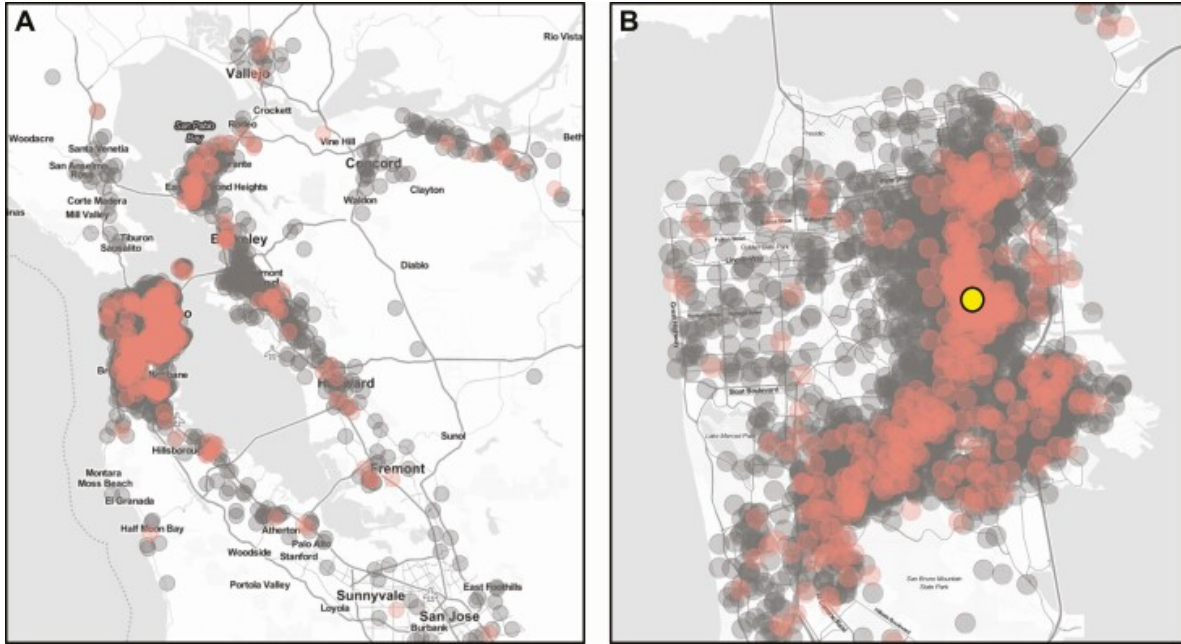
Figure 3.1: Testing catchment area. The location of the 24th & Mission testing site is denoted by the yellow symbol. Negative tests are in gray, and positive tests are shown in red. Household locations shown have a random offset of up to 750 meters to obfuscate the precise addresses of individuals. The testing catchment area encompasses a substantial number of individuals in the surrounding 8 Bay Area counties (*A*). The greatest concentration of individuals reside within San Francisco county (*B*), Map tiles by Stamen Design and data by OpenStreetMap.
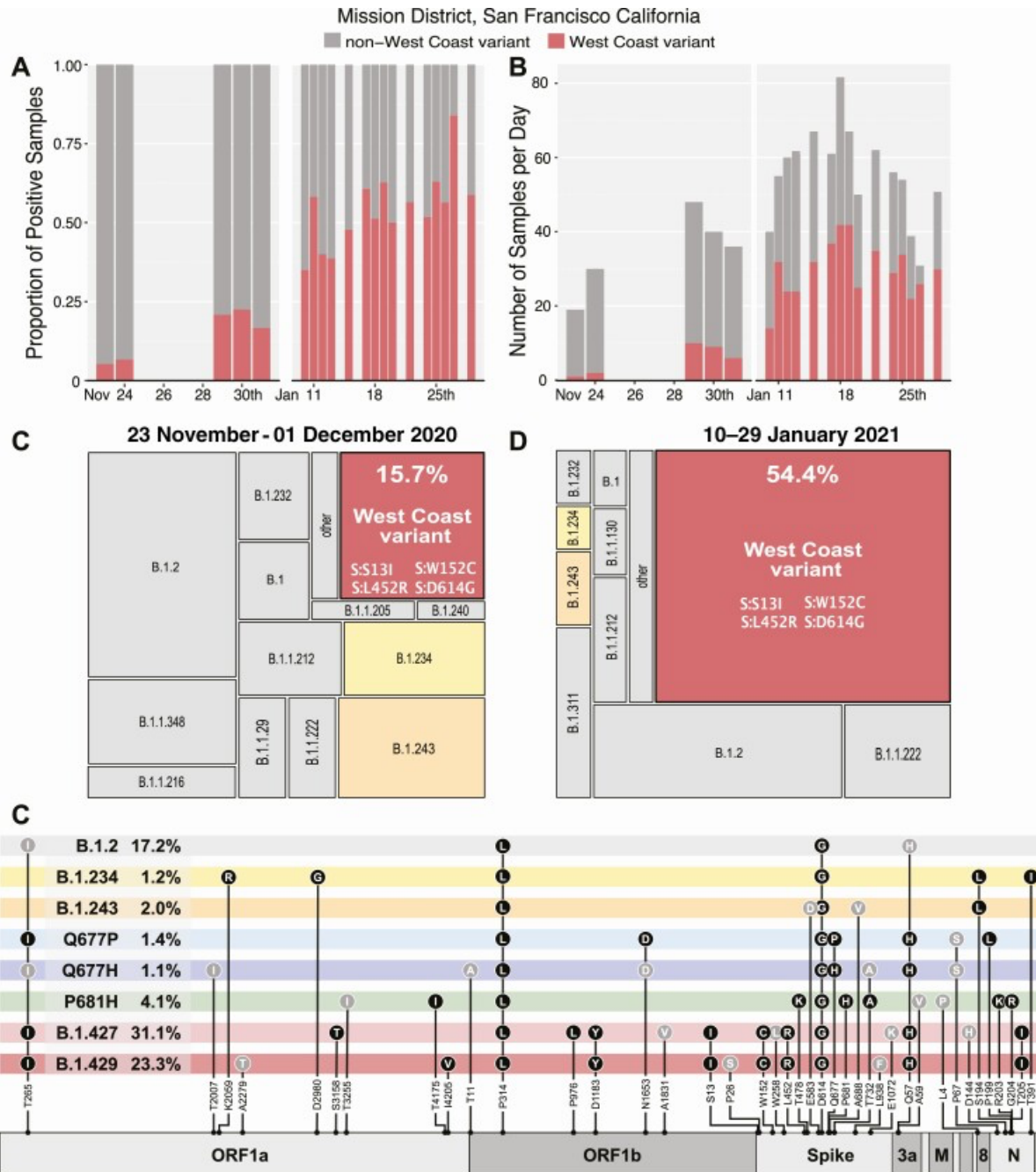
Figure 3.2: Variants observed at 24th & Mission. *A*, Proportion of daily cases belonging to West Coast and non-West Coast variants. *B*, Total number of samples per day. *C*, *D*, Area maps [22] showing the relative proportion of PANGO lineages acquired from full length genomes from the November (N = 191) and January (N = 737) time periods, respectively. *E*, Genome maps for variants detected in this study. Dominant mutations (filled black circles), and nonsynonymous mutations detected at lower frequency in combination with existing lineages (filled gray circles) are shown in gray. Abbreviation: PANGO, Phylogenetic Assignment of Named Global Outbreak.

Table 3.1: Characteristics of Households Included in the Household Attack Rate Analysis, Stratified by Strain.

	Non-West Coast (N = 156)	West Coast			Total (N = 319)
		B.1.427 (N = 90)	B.1.429 (N = 65)	All West Coast (N = 163)[a]	
Race/Ethnicity (most common in household)					
Hispanic/Latinx	143 (91.7%)	78 (86.7%)	62 (95.4%)	146 (89.6%)	289 (90.6%)
Asian	5 (3.2%)	5 (5.6%)	1 (1.5%)	8 (4.9%)	13 (4.1%)
White/Caucasian	4 (2.6%)	3 (3.3%)	0 (0%)	3 (1.8%)	7 (2.2%)
Black or African American	2 (1.3%)	2 (2.2%)	2 (3.1%)	4 (2.5%)	6 (1.9%)
Other	2 (1.3%)	2 (2.2%)	0 (0%)	2 (1.2%)	4 (1.3%)
Has children					
Does not have children	105 (67.3%)	69 (76.7%)	35 (53.8%)	110 (67.5%)	215 (67.4%)
Has children	51 (32.7%)	21 (23.3%)	30 (46.2%)	53 (32.5%)	104 (32.6%)
Location					
San Francisco	118 (75.6%)	71 (78.9%)	39 (60.0%)	115 (70.6%)	233 (73.0%)
Outside San Francisco	38 (24.4%)	19 (21.1%)	26 (40.0%)	48 (29.4%)	86 (27.0%)
Household size					
2 persons	14 (9.0%)	12 (13.3%)	5 (7.7%)	20 (12.3%)	34 (10.7%)
3–4 persons	63 (40.4%)	33 (36.7%)	22 (33.8%)	57 (35.0%)	120 (37.6%)
5+ persons	79 (50.6%)	45 (50.0%)	38 (58.5%)	86 (52.8%)	165 (51.7%)
Household density[b]					
Mean (SD)	1.86 (0.858)	1.91 (0.881)	2.27 (1.22)	2.04 (1.03)	1.95 (0.955)
Median [min, max]	1.88 [0.250, 7.00]	1.67 [0.444, 5.00]	2.00 [0.714, 6.00]	1.69 [0.444, 6.00]	1.75 [0.250, 7.00]

[a]8 households with S gene only sequence available.

[b]Household density missing for 17 households.

Table 3.2: Secondary Household Attack Rates for West Coast Variants, Combined and Disaggregated by B.1.427 and B.1.429.

	Positives Among Tested Contacts (%)	Mean Household Attack Rate (95% CI)	Unadjusted		Adjusted	
			RR (95% CI)	P-value	aRR	P-value
Class						
Non-West Coast	122/415 (29.4%)	25.6% (20.3–31)	…	…	…	…
West Coast	161/452 (35.6%)	35.9% (30.1–41.9)	1.28 (1.00–1.64)	.05	1.25 (.98–1.59)	.07
Lineage						
B.1.427	78/235 (33.2%)	32.9% (25.4–40.6)	1.19 (0.89–1.59)	.20	1.19 (.90–1.59)	.20
B.1.429	79/196 (40.3%)	40.9% (31.5–50.5)	1.43 (1.07–1.91)	.02	1.36 (1.01–1.83)	.04

Relative risks estimated based on Poisson regression using generalized estimating equations and cluster-robust standard errors. Adjustment variables included age group, Latinx/Hispanic race, household size, and household density.

Abbreviations: aRR, adjusted risk ratio; CI, confidence interval; RR, risk ratio.

Table 3.3: Secondary Attack Rate Disaggregated by Covariates.

	Non-West Coast Strain		West Coast Strain	
	Positives Among Tested Contacts (%)	Mean Household Attack Rate (95% CI)	Positives Among Tested Contacts (%)	Mean Household Attack Rate (95% CI)
Location				
San Francisco	88/321 (27.4%)	22.9% (17.2–28.8)	113/316 (35.8%)	37.5% (30.2–44.9)
Outside of San Francisco	34/94 (36.2%)	34% (22.1–46.2)	48/136 (35.3%)	32.1% (22.2–42.1)
Age group				
Age ≤ 12	31/78 (39.7%)	…	41/79 (51.9%)	…
Age > 12	91/337 (27%)	…	120/373 (32.2%)	…
Race/Ethnicity				
Latinx/Hispanic	107/372 (28.8%)	…	136/379 (35.9%)	…
Not Latinx/Hispanic	15/43 (34.9%)	…	25/73 (34.2%)	…
Household size				
2 persons	1/14 (7.1%)	7.1% (0–21.4)	12/20 (60%)	60% (40–80)
3–4 persons	30/115 (26.1%)	26.5% (17.7–35.7)	35/102 (34.3%)	33.3% (23.7–43.6)
5+ persons	91/286 (31.8%)	28.1% (21.1–35.4)	114/330 (34.5%)	32% (25.1–39.2)
Household density				
Bottom half	43/159 (27%)	23.8% (15.7–32.2)	52/176 (29.5%)	33% (24.2–42.3)
Top half	76/243 (31.3%)	27.4% (20.1–35)	101/262 (38.5%)	35.7% (27.9–43.6)

Mean household secondary attack rate only reported disaggregated by household level characteristics.

Abbreviation: CI, confidence interval.

Chapter 3 References

1. Grubaugh ND, Hodcroft EB, Fauver JR, Phelan AL, Cevik M. Public health actions to control new SARS-CoV-2 variants. *Cell* 2021; 184:1127–32.

2. Liu Z, VanBlargan LA, Bloyet L-M, et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host & Microbe* 2021; 29:477– 488.e4.

3. Voloch CM, Francisco R da S, Almeida LGP de, et al. Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. *J Virol* 2021; 95. Available at: https://jvi.asm.org/content/95/10/e00119-21. Accessed 17 May 2021.

4. Weisblum Y, Schmidt F, Zhang F, et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. eLife 2020;9:e61312.

5. Ferrareze PAG, Franceschi VB, Mayer A de M, et al. E484K as an innovative phylogenetic event for viral evolution: genomic analysis of the E484K spike mutation in SARS-CoV-2 lineages from Brazil. *bioRxiv* 2021;2021.01.27.426895.

6. Zeller M, Gangavarapu K, Anderson C, et al. Emergence of an early SARS-CoV-2 epidemic in the United States. medRxiv 2021;2021.02.05.21251235.

7. Fauver JR, Petrone ME, Hodcroft EB, et al. Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States. *Cell* 2020; 181:990–996.e5.

8. Kalinich CC, Jensen CG, Neugebauer P, et al. Real-time public health communication of local SARS-CoV-2 genomic epidemiology. *PLoS Biol* 2020; 18:e3000869.

9. Hodcroft EB, Domman DB, Oguntuyo K, et al. Emergence in late 2020 of multiple lineages of SARS-CoV-2 spike protein variants affecting amino acid position 677. medRxiv 2021;2021.02.12.21251658.

10. COVID Testing Pop-up site Established at 24th & Mission to Help Combat Surge in the City's Most Impacted Community During this Holiday Season. 2020. Available at: https://www.sfdph.org/dph/alerts/files/12.20.20_CCSF_Press_Release.pdf.

11. Pilarowski G, Marquez C, Rubio L, et al. Field performance and public health response using the BinaxNOWTM rapid severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Antigen Detection Assay During Community-Based Testing. *Clin Infect Dis* 2021; 73:e3098–101.

12. Pilarowski G, Lebel P, Sunshine S, et al. Performance characteristics of a rapid severe acute respiratory syndrome coronavirus 2 antigen detection assay at a public plaza testing site in San Francisco. *J Infect Dis* 2021; 223:1139–1144.

13. Rubio LA, Peng J, Rojas S, et al. The COVID-19 symptom to isolation cascade in a Latinx community: a call to action. *Open Forum Infect Dis* 2021; 8:ofab023.

14. Kerkhoff AD, Sachdev D, Mizany S, et al. Evaluation of a novel community-based COVID-19 'test-to-care' model for low-income populations. *PLoS One* 2020; 15:e0239400.

15. ARTIC Project . hCoV-2019/nCoV-2019 version 3 amplicon set. Available at: https://artic.network/resources/ncov/ncov-amplicon-v3.pdf. Accessed 25 February 2021.

16. Kalantar KL, Carvalho T, de Bourcy CFA, et al. IDseq—an open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. *GigaScience* 2020; 9. doi: 10.1093/gigascience/giaa111.

17. Gelman A, Hill J. *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press, 2007.

18. Laan MJ van der, Rose S. *Targeted learning in data science: causal inference for complex longitudinal studies*. London: Springer International Publishing, 2018. Available at: https://www.springer.com/gp/book/9783319653037. Accessed 24 February 2021.

19. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* 2017; 1:33–46.

20. Chamie G, Marquez C, Crawford E, et al. Community transmission of severe acute respiratory syndrome coronavirus 2 disproportionately affects the Latinx population during Shelter-in-Place in San Francisco. *Clin Infect Dis* 2021; 73(Suppl. 2):S127–35.

21. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018; 34:4121–3.

22. Laserson U. laserson/squarify. 2021. Available at: https://github.com/laserson/squarify. Accessed 26 February 2021.

23. Zhang W, Davis BD, Chen SS, Martinez JMS, Plummer JT, Vail E. Emergence of a novel SARS-CoV-2 strain in Southern California, USA. medRxiv 2021;2021.01.18.21249786.

24. Hoffmann M, Kleine-Weber H, Pöhlmann S.A Multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol Cell* 2020; 78:779–784.e5.

25. Örd M, Faustova I, Loog M. The sequence at spike S1/S2 site enables cleavage by furin and phospho-regulation in SARS-CoV2 but not in SARS-CoV1 or MERS-CoV. *Sci Rep* 2020; 10:16944.

26. Johnson BA, Xie X, Bailey AL, et al. Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature* 2021; 591:293–299.

27. Jaimes JA, Millet JK, Whittaker GR. Proteolytic cleavage of the SARS-CoV-2 spike protein and the role of the novel S1/S2 site. *iScience* 2020; 23:101212.

28. Madewell ZJ, Yang Y, Longini IM Jr, Halloran ME, Dean NE. Household Transmission of SARS-CoV-2: a systematic review and meta-analysis. *JAMA Netw Open* 2020; 3:e2031756.

29. Volz E, Mishra S, Chand M, et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* 2021; 593:266–269.

Chapter 4: Characterization of SARS-CoV-2 transmission and immune response dynamics

in a vaccinated household

This chapter is a reprint of the following publication:

Liu J, Laurie MT, Rubio L, Vazquez SE, Sunshine S, Mitchell AM, Hapte-Selassie M, Mann SA, Pilarowski G, Black D, Marquez C, Rojas S, Lionakis MS, Petersen M, Whitman JD, Jain V, Anderson M, Havlir D, DeRisi J. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Transmission Dynamics and Immune Responses in a Household of Vaccinated Persons. *Clin Infect Dis*. 2022 Aug 24;75(1):e303-e306. doi: 10.1093/cid/ciac029. PMID: 35037050; PMCID: PMC8807302.

Supplemental files are included with original published work.

Abstract

While SARS-CoV-2 vaccines prevent severe disease effectively, postvaccination "breakthrough" COVID-19 infections and transmission among vaccinated individuals remain ongoing concerns. We present an in-depth characterization of transmission and immunity among vaccinated individuals in a household, revealing complex dynamics and unappreciated comorbidities, including autoimmunity to type 1 interferon in the presumptive index case.

Main Text

Coronavirus disease 2019 (COVID-19) has caused over 230 million cases of infection worldwide, leading to more than 4.7 million deaths due to COVID-19 [1]. Global vaccination efforts have so far administered 6.1 billion vaccine doses [2]. In the United States, 3 Food and Drug Administration (FDA)–authorized vaccines have been widely distributed: BNT162b2 by Pfizer/BioNTech, mRNA-1273 by Moderna, and JNJ-78436735 by Johnson & Johnson/Janssen. Each has demonstrated, through clinical trials and retrospective studies, the capacity to prevent symptomatic infection and severe disease [3].

Approximately 50% of the US population is considered fully vaccinated. Many households have mixed populations of adults and children with variable completion of COVID-19 vaccination [2]. Furthermore, most severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) lineages have been outcompeted and replaced by newer variants of concern, including the Delta and Gamma variants. Further, many spike protein mutations associated with neutralizing antibody escape (K417N/T, R346K, L452R, T478K, E484K/Q, N501Y) have emerged [4, 5]. Given these factors, COVID-19 infections in fully vaccinated people (ie, breakthrough) are well documented [6]. However, there have been relatively few detailed studies to date of household transmission

trajectories, especially in households with individuals who received different vaccines, or who have different vaccine completion statuses.

Here, we describe a household cluster of Gamma variant COVID-19 cases occurring in vaccinated family members living in co-residence that resulted in mixed clinical outcomes. A detailed inspection of the epidemiological and clinical features of these cases, together with serology testing and genomic sequencing, suggest complex factors including partial immunity and unrecognized underlying autoimmunity, as potential contributors to breakthrough infections. Our data add to the rapidly emerging literature on SARS-CoV-2 transmission dynamics within households of vaccinated persons.

Methods

Description of Individuals in the Study Household

Individuals 1–5 lived together in the same residence, where they ate, slept, and socialized with one another in an unmasked setting. Individual 6 lived separately but frequented the home of Individuals 1–5. Together, these individuals also attended weekly community events, such as religious services, together as 1 large group. Each individual was thus exposed to one another either through co-residence or frequent visitation.

Individual 1 is an 80-year-old man with diabetes and asthma who received the BNT162b2/Pfizer vaccine on 20 April and 10 May 2021. On 13 May, malaise, myalgia, and diarrhea developed. On 19 May, a SARS-CoV-2 polymerase chain reaction (PCR) test was positive, and on 20 May, he presented to a local hospital, had hypoxia, and was admitted for inpatient management. Due to severe COVID-19, acute respiratory distress syndrome (ARDS), and respiratory failure, he required mechanical ventilation. He received remdesivir,

dexamethasone, and tociluzimab and improved, was weaned from the ventilator, and was discharged home on 2 June.

Individual 2 is a 36-year-old woman who received the JNJ-78436735/Janssen vaccine on 10 April 2021. On 16 May, she had onset of fever, cough, rhinorrhea, and headache. On 19 May, a PCR test was positive. On 23 May, a BinaxNOW (Abbott) rapid antigen test was positive. She did not require care at a health facility and improved with self-monitoring at home.

Individual 3 is a 60-year-old woman who received the mRNA-1273/Moderna vaccine on 9 March and 6 April 2021. On 19 May, she had onset of fever, chills, cough, and rhinorrhea. On 20 May, a SARS-CoV-2 PCR test was positive, and on 23 May, a BinaxNOW test was positive. She also did not require care at a health facility and improved with self-monitoring at home.

Individual 4 is an 84-year-old woman who received the mRNA-1273/Moderna vaccine on 25 February and 26 March 2021. After members of her family tested positive for COVID-19, she began home-based quarantine on 20 May. On 23 May, a BinaxNOW test was negative.

Individual 5 is a 40-year-old man who had tested positive for SARS-CoV-2 the previous year on 24 July 2020. At that time, he isolated with Individual 6. Individual 5 received the JNJ-78436735/Janssen vaccine on 10 April 2021. Although he did not quarantine separately from family members who tested positive, a SARS-CoV-2 PCR test on 22 May was negative.

Individual 6 is a 60-year-old woman who directly cared for Individual 5 when he tested positive for SARS-CoV-2 in July 2020. Despite being unable to quarantine, she tested negative for SARS-CoV-2 and did not develop any COVID-like symptoms. On 17 May 2021, she received the first dose of BNT162b2/Pfizer vaccine. Although she lived apart from Individuals 1–5, she visited their home frequently and attended community events with them. When her BinaxNOW test was negative on 23 May, she had not yet received a second dose of the vaccine.

Timelines of vaccination, COVID-19 symptom onset, and testing history are summarized in Figure 4.1A and Supplementary Table 1.

Results

SARS-CoV-2 positivity as determined by quantitative PCR (qPCR) amplification of the nasal swab samples corroborated the BinaxNOW results for each household member. Viral genome sequences were recovered from the 3 individuals who tested positive. Sequences consistent with the Gamma variant were recovered from Individual 2 (90% genome coverage; GISAID: EPI_ISL_2508365) and Individual 3 (98% genome coverage; GISAID: EPI_ISL_2508366) (Figure 4.2, BioProject PRJNA790937). Despite incomplete recovery, the partial sequence from Individual 1 (17%) contained mutations consistent with the Gamma variant (Supplementary Table 2). Characteristic mutations of concern (K417T, E484K, and N501Y) were observed [4, 5]. Analysis of the consensus genomes from Individuals 2 and 3 revealed only a single nucleotide difference (G17122T, leading to a ORF1b:A1219S amino acid substitution).

Serum samples from the 5 household members were analyzed for SARS-CoV-2 neutralizing antibodies using a pseudo-virus neutralization assay [9]. Sera from members of this household demonstrated a wide range of neutralization (Figure 4.1B). Individual 1 had a much lower neutralizing antibody titer compared with the fully vaccinated individuals (D614G 50% neutralization titer $[NT_{50}]$ = 4.4× lower, Gamma NT_{50} = 6.3× lower), despite being measured 14 days post–symptom onset and 17 days after his second vaccine dose. Conversely, despite only partial vaccination, Individual 6 had a very high neutralizing antibody titer (D614G NT_{50} = 4.5× higher, Gamma NT_{50} = 5.0× higher) versus the healthy vaccinated cohort. Although this may have been related to caring for Individual 5 a year prior, Individual 6 had negative serology on the anti–

SARS-CoV-2-N immunoglobulin G (IgG) Abbott Architect test. Finally, while Individuals 2, 3, and 4 had neutralizing antibody titers in the typical range of fully vaccinated individuals, Individuals 2 and 3 ultimately tested positive for COVID-19. Taken together, our observations indicate that fully vaccinated individuals may be at risk of breakthrough infection when living in households with sustained close contact with infected individuals.

The neutralization efficacy of patients' sera against the Gamma variant pseudo-type was approximately 2-fold lower than the measured NT_{50} against wild-type virus (D614G spike mutation only). This observation is consistent with previously described decreases in neutralization against variants, especially those harboring mutations at E484K [4, 5, 7].

Additionally, we tested for anti–interferon (IFN)-α2 autoantibodies, a marker correlated with severe COVID-19 and poor patient outcomes [10, 11]. Using serum from patients with autoimmune polyglandular syndrome type 1 (APS1), an autoimmune syndrome where patients frequently develop an abundance of anti–IFN-α2 antibodies, as a benchmark for verified IFN autoimmunity, we measured for anti–IFN-α2 antibody presence using a radioligand binding assay (RLBA) [10]. Serum from Individual 1, who had the most severe response to infection, exhibited positive anti–IFN-α2 antibody signal while the other family members had negative titers (Figure 4.1C).

Discussion

We describe a family of mixed vaccination statuses who experienced various clinical trajectories after a Gamma variant COVID-19 exposure in the household. Although coverage of the recovered SARS-CoV-2 genome from Individual 1 is incomplete, and Individuals 2 and 3 differ by 1 amino acid substitution, the rarity of the Gamma variant (6.5% of all sequences

submitted to GISAID from San Francisco County from April to June) supports the conjecture that infection of this household is derived from a common source. Furthermore, all other Gamma variant sequences from this time period had 3–32 (mean = 13, median = 14) nucleotide substitutions compared with this household, strongly suggesting direct transmission between household individuals as opposed to coincidental, simultaneous infection outside the home.

Clinical trajectories experienced by household individuals ranged from severe illness requiring hospitalization, to mild symptomatic illness, to avoiding COVID-19 infection altogether. Individual 1, who had low titers of neutralizing antibodies following vaccination, still developed severe COVID-19 infection. Testing for anti–IFN-α2 autoantibodies revealed that serum from Individual 1 contained high levels of antibodies against IFN-α2, a trait enriched among patients with life-threatening COVID-19 pneumonia [11]. Although the presence of such autoantibodies can be clinically silent, they appear to play an influential role in patient outcomes for SARS-CoV-2 infection [12].

Comorbidities such as autoimmune disease caused by anti-IFN autoantibodies can lead to decreased protection against circulating variants with spike mutations conferring neutralization escape and thus raise the risk of breakthrough infections [11]. With household exposure to COVID-19, even fully vaccinated individuals with typical levels of neutralizing antibodies are at risk of infection. These data are strongly consistent with intrahousehold transmission among 3 vaccinated household members in this study, and these data highlight the inherent complexities of individuals, including unrealized underlying autoimmunity, that may contribute to transmission dynamics. These data support the urgency for continued vaccination, boosters, and next-generation vaccines that contain mutations known to confer immune escape potential.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.
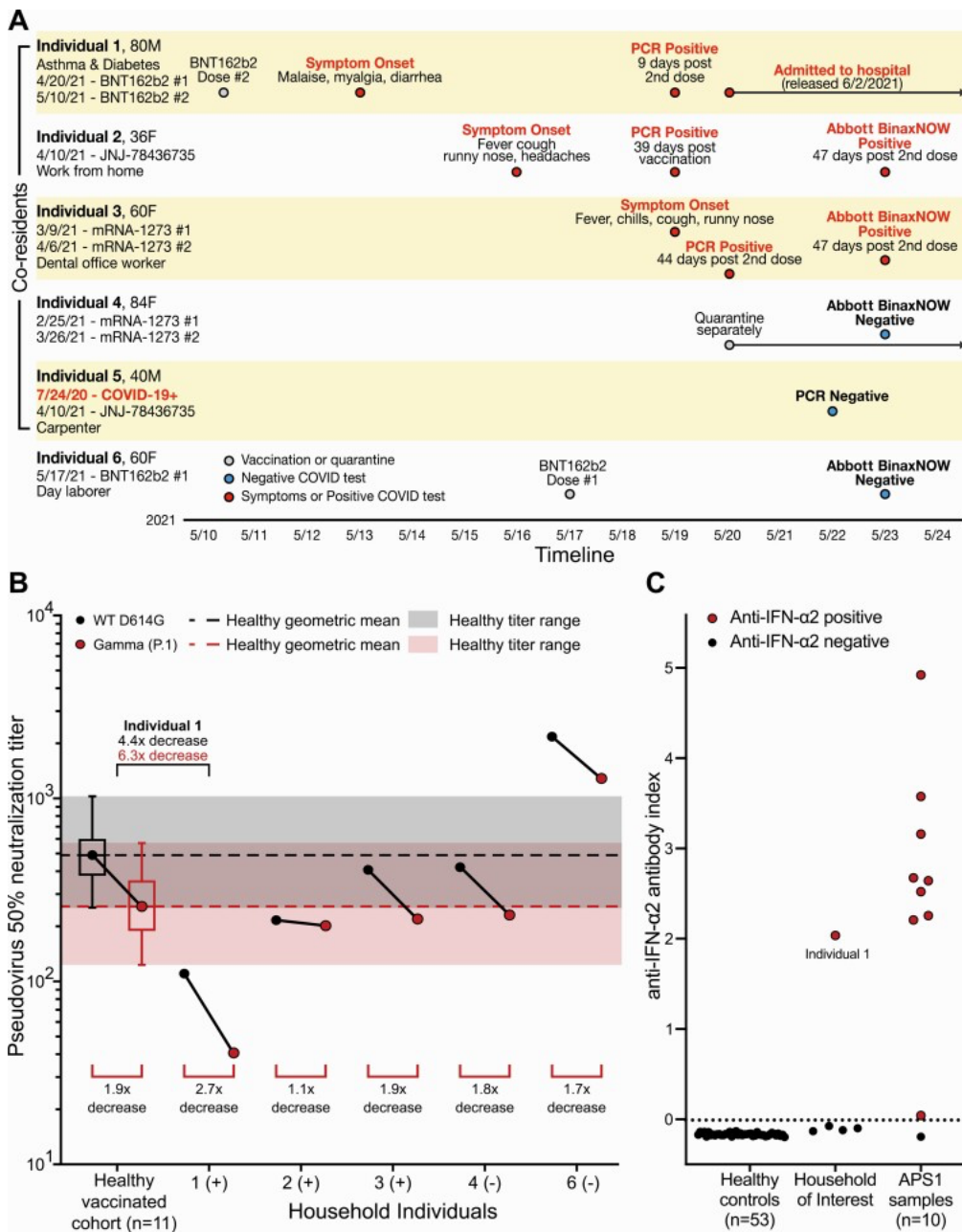
Figure 4.1. Serum samples from household individuals reveal diverse neutralization capabilities as well as presence of anti–IFN-α2 auto-antibodies in Individual 1. (A) Timeline illustrating the order of events experienced by individuals in the study household, including vaccination, symptom onset, and test results. Additional details are available in Supplementary Table 1. (B) Plot of 50% pseudo-virus neutralization titers (NT$_{50}$) of serum samples from healthy vaccinated controls (n = 11) collected 12–60 days post–second dose (average = 26.4 days; details of serum collection timing relative to vaccination and positive COVID-19 tests are described in Supplementary Table 3). For the healthy vaccinated donor cohort, geometric mean titer (dashed lines), interquartile range (boxes), and full range (shaded region) are shown for D614G (black)

51

and Gamma (red) pseudo-viruses. NT$_{50}$ values for Gamma variant pseudo-virus were approximately 2-fold lower than D614G pseudo-virus for the healthy vaccinated cohort and most household members sera, apart from Individual 2. All household member serum neutralization titers were within or above the range of healthy donor titers, except for Individual 1, whose neutralization titers for D614G and Gamma were 4.4-fold and 6.3-fold lower than those in healthy controls, respectively. (C) Detection by radioligand binding assay reveals that anti–IFN-α2 autoantibodies are absent from all assayed prepandemic healthy controls (n = 42) and vaccinated healthy controls (n = 11) [7]. In this household, only Individual 1 demonstrated the presence of anti–IFN-α2 auto-antibodies. Autoimmune polyglandular syndrome type 1 (APS1) patient sera are used as positive controls [8]; negative controls are from pre-COVID healthy blood donor plasma or the healthy vaccinated donor cohort. Abbreviations: COVID-19, coronavirus disease 2019; F, female; IFN, interferon; M, male; PCR, polymerase chain reaction.
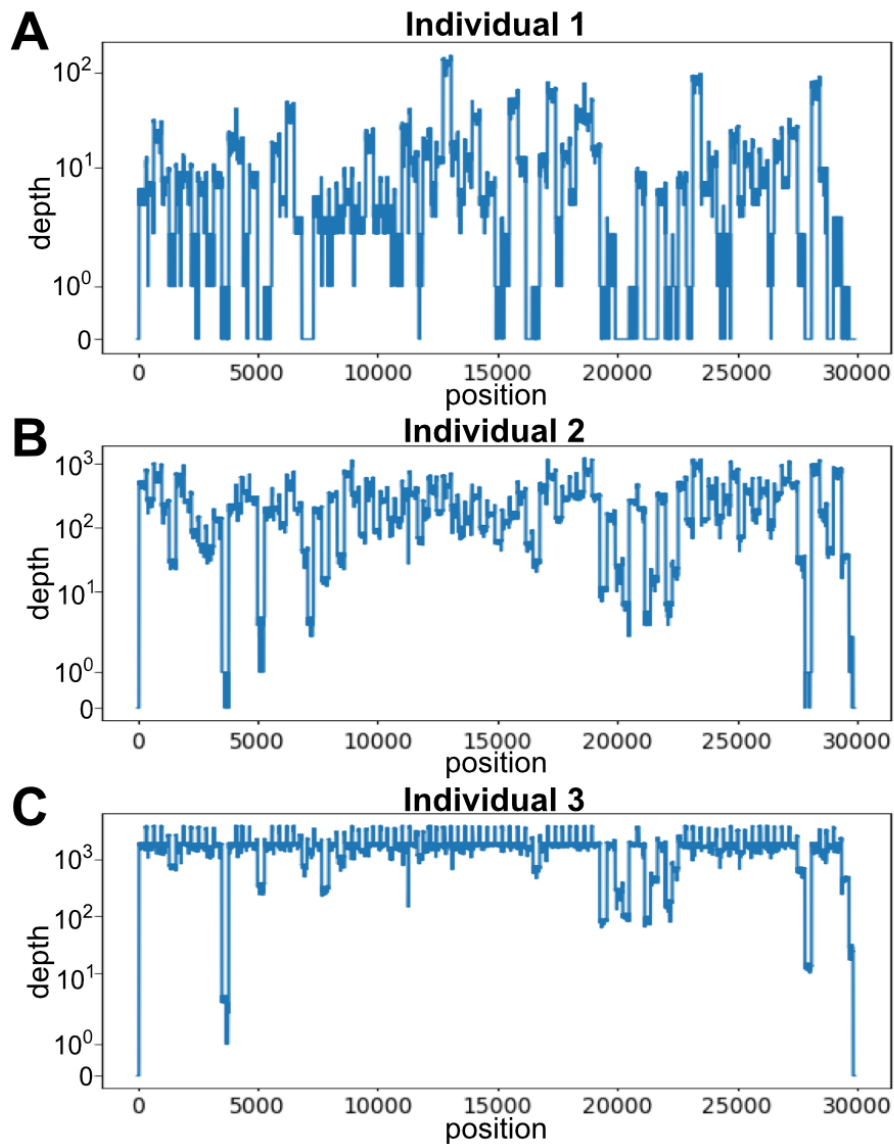
Figure 4.2. Sequencing depth and coverage of recovered SARS-CoV-2 genomes. Genome depth and coverage of SARS-CoV-2 Gamma variant recovered from (A) Individual 1, (B) Individual 2, and (C) Individual 3.

Chapter 4 References

1. Johns Hopkins University. *COVID-19 dashboard*. Available at: https://coronavirus.jhu.edu/map.html. Accessed 11 September 2021.

2. Ritchie H, Mathieu E, Rodés-Guirao L, et al. Our world in data: coronavirus pandemic (COVID-19). Our world in data 2020. Available at: https://ourworldindata.org/covid-vaccinations. Accessed 11 September 2021.

3. Moline HL, Whitaker M, Deng L, et al. Effectiveness of COVID-19 vaccines in preventing hospitalization among adults aged ≥65 years—COVID-NET, 13 states, February-April 2021. *MMWR Morb Mortal Wkly Rep* 2021; 70:1088–93.

4. Garcia-Beltran WF, Lam EC, St Denis K, et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* 2021; 184:2372–2383.e9.

5. Wang Z, Schmidt F, Weisblum Y, et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* 2021; 592:616–22.

6. Vignier N, Bérot V, Bonnave N, et al. Breakthrough infections of SARS-CoV-2 gamma variant in fully vaccinated gold miners, French Guiana, 2021. *Emerg Infect Dis* 2021; 27:2673–6.

7. Laurie MT, Liu J, Sunshine S, et al. SARS-CoV-2 variant exposures elicit antibody responses with differential cross-neutralization of established and emerging strains including Delta and Omicron. *J Infect Dis* 2022; jiab635.

8. Ferre EMN, Rose SR, Rosenzweig SD, et al. Redefined clinical features and diagnostic criteria in autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy. *JCI Insight* 2016; 1:e88782.

9. Hoffmann M, Kleine-Weber H, Pöhlmann S. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol Cell* 2020; 78:779–784, e5.

10. van der Wijst MGP, Vazquez SE, Hartoularos GC, et al. Type I interferon autoantibodies are associated with systemic immune alterations in patients with COVID-19. *Sci Transl Med* 2021; 13:eabh2624.

11. Bastard P, Rosen LB, Zhang Q, et al. Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science* 2020; 370:eabd4585.

12. Bastard P, Gervais A, Le Voyer T, et al. Autoantibodies neutralizing type I IFNs are present in ~4% of uninfected individuals over 70 years old and account for ~20% of COVID-19 deaths. *Sci Immunol* 2021; 6:eabl4340.

Chapter 5: Interrogation of cross-neutralization of SARS-CoV-2 variants by sera from infected and vaccinated populations

This chapter is a reprint of the following publication:

Laurie MT, Liu J, Sunshine S, Peng J, Black D, Mitchell AM, Mann SA, Pilarowski G, Zorn KC, Rubio L, Bravo S, Marquez C, Sabatino JJ, Mittl K, Petersen M, Havlir D, DeRisi J. SARS-CoV-2 Variant Exposures Elicit Antibody Responses With Differential Cross-Neutralization of Established and Emerging Strains Including Delta and Omicron. J Infect Dis. 2022 Jun 1;225(11):1909-1914. doi: 10.1093/infdis/jiab635. PMID: 34979030; PMCID: PMC8755395.

Supplemental files are included with original published work.

Abstract

The wide spectrum of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants with phenotypes impacting transmission and antibody sensitivity necessitates investigation of immune responses to different spike protein versions. Here, we compare neutralization of variants of concern, including B.1.617.2 (delta) and B.1.1.529 (omicron), in sera from individuals exposed to variant infection, vaccination, or both. We demonstrate that neutralizing antibody responses are strongest against variants sharing certain spike mutations with the immunizing exposure, and exposure to multiple spike variants increases breadth of variant cross-neutralization. These findings contribute to understanding relationships between exposures and antibody responses and may inform booster vaccination strategies.

Main Text

Genomic surveillance of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) continues to identify a diverse spectrum of emerging variants possessing mutations in the spike gene, the main viral determinant of cellular entry and primary target of neutralizing antibodies [1]. Many spike mutations likely result from selective pressure that improves viral fitness through increased transmissibility or evasion of host immunity [2, 3]. Studies have demonstrated that sera from vaccinated and naturally infected individuals yield diminished neutralizing activity against certain variants, including the globally dominant delta variant [4]. Because serum neutralization titer is an important correlate of real-world protective immunity, these findings suggest that antibody responses elicited by exposure to ancestral spike versions (Wuhan or D614G) will be less effective at preventing future infection by certain variants [5]. However, the diversity and prevalence of variants have fluctuated greatly throughout the pandemic, creating a complex

57

population of individuals that may have inherently different capacity to neutralize certain variants depending on the specific genotype of their previous exposures, including vaccination [6].

In this study, we address the question of variant-elicited immune specificity by determining the breadth of neutralizing activity elicited by exposure to specific SARS-CoV-2 variants, vaccines, or both. To accomplish this, we collected serum from subjects with prior infections by variants B.1 (D614G mutation only), B.1.429 (epsilon), P.2 (zeta), B.1.1.519, and B.1.617.2 (delta), which were identified by viral sequencing. We also collected serum from mRNA vaccine recipients who were infected with the B.1 ancestral spike lineage prior to vaccination, infected with B.1.429 prior to vaccination, or had no prior infection. We measured and compared the neutralization titer of each serum cohort against a panel of pseudoviruses representing each different exposure variant plus the variants of concern B.1.351 (beta), P.1 (gamma), B.1.617, B.1.617.2 (delta), and B.1.1.529 (omicron), which have 1 or more spike mutations of interest in common with 1 of the exposure variants. Our results provide a quantitative comparison of the degree of neutralization specificity produced by different exposures. We also demonstrate the effect of serial exposure to different spike versions in broadening the cross-reactivity of neutralizing antibody responses. Together, these findings describe correlates of protective immunity within the rapidly evolving landscape of SARS-CoV-2 variants and are highly relevant to the design of future vaccination strategies targeting spike antigens.

Methods

Serum Collection

Samples for laboratory studies were obtained under informed consent from participants in an ongoing community program Unidos en Salud, which provides SARS-CoV-2 testing, genomic

surveillance, and vaccination services in San Francisco, California [7]. Subjects with and without symptoms of coronavirus disease 2019 (COVID-19) were screened with the BinaxNOW rapid antigen assay (supplied by California Department of Public Health). Positive rapid tests were followed by immediate disclosure and outreach to household members for testing, supportive community services, and academic partnership for research studies. All samples were sequenced using ARTIC Network V3 primers on an Illumina NovaSeq platform and consensus genomes generated from the resulting raw.fastq files using IDseq [8].

Convalescent serum donors were selected based on sequence-confirmed infection with the following variants of interest: B.1 (D614G mutation only; n = 10 donors), B.1.429 (epsilon; n = 15), B.1.1.519 (n = 6), P.2 (zeta; n = 1), B.1.526 (iota; n = 1), B.1.617.2 (delta; n = 3), D614G infection with subsequent BNT162b2 vaccination (n = 8), and B.1.429 infection with subsequent BNT162b2 vaccination (n = 17). Serum was also collected from healthy recipients of 2 (n = 11) or 3 (n = 7) doses of BNT162b2 or mRNA-1273 vaccines who were confirmed to have no prior SARS-CoV-2 infection by anti–SARS-CoV-2 nucleocapsid IgG assay [9]. All serum was collected from donors an average of 34 days (standard deviation 16.6 days) after exposure to either SARS-CoV-2 or the most recent dose of mRNA vaccine. For pooled serum experiments, samples from the same exposure group were pooled at equal volumes. Serum samples from the closely related exposures P.2 and B.1.526 were pooled together for the E484K exposure pool, and samples from BNT162b2 and mRNA-1273 exposures were pooled together for the vaccine exposure pool because of the very similar neutralization specificity observed in individual tests of these sera. Serum samples were heat inactivated at 56°C for 30 minutes prior to experimentation. Relevant serum sample metadata and exposure grouping is shown in Supplementary Table 1.

Pseudovirus Production

SARS-CoV-2 pseudoviruses bearing spike proteins of variants of interest were generated using a recombinant vesicular stomatitis virus (VSV) expressing green fluorescent protein (GFP) in place of the VSV glycoprotein (rVSV Δ G-GFP) described previously [10]. The following mutations representative of specific spike variants were cloned in a cytomegalovirus enhancer-driven expression vector and used to produce SARS-CoV-2 spike pseudoviruses: B.1 (D614G), B.1.429/epsilon (S13I, W152C, L452R, D614G), P.2/zeta (E484K, D614G), B.1.351/beta (D80A, D215G, Δ242-244, K417N, E484K, N501Y, D614G, A701V), P.1/gamma (L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I, V1176F), B.1.1.519 (T478K, D614G, P681H, T732A), B.1.617 (L452R, E484Q, D614G, P681R), B.1.617.2/delta (T19R, T95I, G142D, Δ157-158, L452R, T478K, P681R, D614G, D950N), and B.1.1.529/omicron (32 spike mutations). All pseudovirus spike mutations are listed in Supplementary Table 2. Pseudoviruses were titered on Huh7.5.1 cells overexpressing ACE2 and TMPRSS2 (gift of Andreas Puschnik) using GFP expression to measure the concentration of focus forming units (ffu).

Pseudovirus Neutralization Experiments

Huh7.5.1-ACE2-TMPRSS2 cells were seeded in 96-well plates at a density of 7000 cells/well 1 day prior to pseudovirus inoculation. Serum samples were diluted into complete culture media (Dulbecco's Modified Eagle's Medium with 10% fetal bovine serum, 10mM HEPES, 1 × Pen-Strep-Glutamine) using the LabCyte Echo 525 liquid handler and 1500 ffu of each pseudovirus was added to the diluted serum to reach final dilutions ranging from 1:40 to 1:5120, including no-serum and no-pseudovirus controls. Serum/pseudovirus mixtures were incubated at 37°C for 1 hour before being added directly to cells. Cells inoculated with

serum/pseudovirus mixtures were incubated at 37°C and 5% CO_2 for 24 hours, resuspended using 10 x TrypLE Select (Gibco), and cells were assessed with the BD Celesta flow cytometer. The World Health Organization International Reference Standard 20/150 was used to validate the pseudovirus assay and compare serum neutralization titers (Supplementary Table 3) [11]. All neutralization assays were repeated in a total of 3 independent experiments with each experiment containing 2 technical replicates for each condition. Cells were verified to be free of mycoplasma contamination with the MycoAlert Mycoplasma detection kit (Lonza).

Data Analysis

Pseudovirus flow cytometry data were analyzed with FlowJo to determine the percentage of GFP-positive cells, indicating pseudovirus transduction. Percent neutralization for each condition was calculated by normalizing GFP-positive cell percentage to no-serum control wells. The 50% and 90% neutralization titers (NT_{50} and NT_{90}) were calculated from 8-point response curves generated in GraphPad Prism 7 using 4-parameter logistic regression. The fold-change in pseudovirus neutralization titer in each serum group was calculated by normalizing each variant NT_{50} and NT_{90} value to D614G pseudovirus NT_{50} and NT_{90} values in the same serum group. To compare neutralization titer across a panel of different pseudoviruses and serum groups, the $\log_2$ fold-change compared to D614G pseudovirus was reported.

Results

We compared NT_{50} and NT_{90} of D614G and B.1.429 (epsilon) pseudoviruses in individual serum samples from subjects exposed to D614G infection, B.1.429 infection, mRNA vaccination, D614G infection with subsequent mRNA vaccination, and B.1.429 infection with subsequent

mRNA vaccination (Figure 5.1). Fold-changes in both NT_{50} and NT_{90} are reported because these values often differ in magnitude due to differences in neutralization curve slope between different variants and sera. In D614G-exposed and vaccine-exposed serum, we observed approximately 2- to 3-fold decreases in average neutralization titer against B.1.429 pseudovirus compared to D614G pseudovirus. As expected, B.1.429-exposed serum neutralized B.1.429 pseudovirus more efficiently than D614G pseudovirus. Of note, previous infection with either D614G or B.1.429 followed by vaccination led to substantially higher neutralization titers against both pseudoviruses. In contrast to other exposure groups, serum from vaccine recipients previously infected by B.1.429 neutralized D614G and B.1.429 at similar titers, with only a 1.3-fold difference in NT_{90}, indicating that exposure to multiple spike variants elicits a potent response with specificity toward the breadth of prior exposures.

We next investigated how exposure impacts neutralization specificity by crossing a panel of 8 different spike variants against serum pools elicited by 9 different prior exposures (Figure 5.2 and Supplementary Table 3). A range of reductions in neutralization titer relative to D614G pseudovirus were observed, with B.1.617.2 (delta), B.1.351 (beta), and B.1.1.529 (omicron) exhibiting the greatest resistance to neutralization in serum from vaccinated or D614G-exposed individuals with up to 4-fold, 12-fold, and 65-fold reductions in NT_{90}, respectively. However, for most variants, reductions in neutralization titer were considerably smaller or absent in serum from subjects previously exposed to a variant bearing some or all of the same spike mutations as the variant being tested. Specifically, prior exposure to the E484K mutation in the spike receptor binding domain (RBD) produced the greatest neutralization of 4 tested variants with mutations at the E484 position: B.1.617, P.1 (gamma), P.2 (zeta), and B.1.351 (beta). Similarly, B.1.617.2 (delta) was neutralized more effectively by serum elicited by partially homologous exposures

B.1.1.519 and B.1.429, and was neutralized most effectively by serum elicited by fully homologous B.1.617.2 exposure. Conversely, in B.1.617.2-exposed serum we observed the least efficient neutralization of the highly divergent spike variants P.1 and B.1.351. Interestingly, although B.1.1.529 (omicron) substantially escaped neutralization in all convalescent sera and serum from recipients of 2 vaccine doses, a much more modest 4- to 8-fold reduction in neutralization titer was observed in sera from individuals with previous infection plus vaccination or three vaccine doses.

Discussion

In this study, we observe that vaccination and natural SARS-CoV-2 infection elicit neutralizing antibody responses that are most potent against variants that bear spike mutations present in the immunizing exposure. This trend is exemplified by variants with mutations at the spike E484 position, which were neutralized more effectively by E484K-exposed serum than other serum types. Importantly, we also show that B.1.617.2 (delta) is neutralized more effectively by serum elicited by prior exposure to 3 different variants—B.1.429, B.1.1.519, and B.1.617.2— which have separate sets of spike mutations partially or fully overlapping with mutations in B.1.617.2. These effects are presumably due to the shared L452R RBD mutation in B.1.429 and B.1.617.2, and the shared T478K RBD mutation and P681 furin cleavage site mutation found in both B.1.1.519 and B.1.617.2. The poor neutralization of P.1 and B.1.351 by delta-exposed serum further reinforces the notion that cross-neutralization is heavily impacted by antigenic distance between variants [12]. Together, these results demonstrate that serum neutralization specificity is strongest against variants fully homologous to the exposure, but even single shared spike

63

mutations, particularly those in highly antigenic regions such as the RBD, can enhance cross-neutralization as supported in other studies [3, 6, 13].

This study also demonstrates the effect of serial exposure to repeated or novel versions of spike on neutralizing antibody response. Infection with B.1.429 (epsilon) followed by vaccination led to greater cross-neutralization of B.1.429 and B.1.617.2 (delta) compared to vaccination alone or D614G infection plus vaccination, supporting the notion that exposure to multiple spike variants expands neutralization breadth. Repeated immunizing exposures from infection plus vaccination or booster vaccination led to both an overall increase in neutralization titers and generally broadened neutralization specificity, particularly towards B.1.1.529 (omicron), which was neutralized most effectively by serum from recipients of 3 vaccine doses. A limitation of this study is the relatively small number of serum samples; however, the shift in neutralization titer between D614G and variant pseudoviruses shows strong consistency between samples.

These serology data leverage human exposures to an array of naturally occurring spike mutations, including those relevant to the globally dominant B.1.617.2 and recently ascendant B.1.1.529 variants, providing a real-world complement to previous animal studies investigating heterologous boosting or multivalent vaccination strategies [14, 15]. Our findings suggest that immunity acquired through natural infection will differ significantly between populations in different regions of the world due to highly variable prevalence of different SARS-CoV-2 variants throughout the course of the ongoing pandemic. These results also reinforce the urgent need for widespread booster vaccination and contribute additional evidence suggesting that heterologous or multivalent boosting strategies may be important and effective measures to address newly emergent variants such as the highly immune evasive B.1.1.529 (omicron). Future studies investigating immune

responses to additional emerging variants in vaccinated and unvaccinated individuals will contribute to identifying spike antigen versions that elicit broadly neutralizing antibody responses.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online. Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.
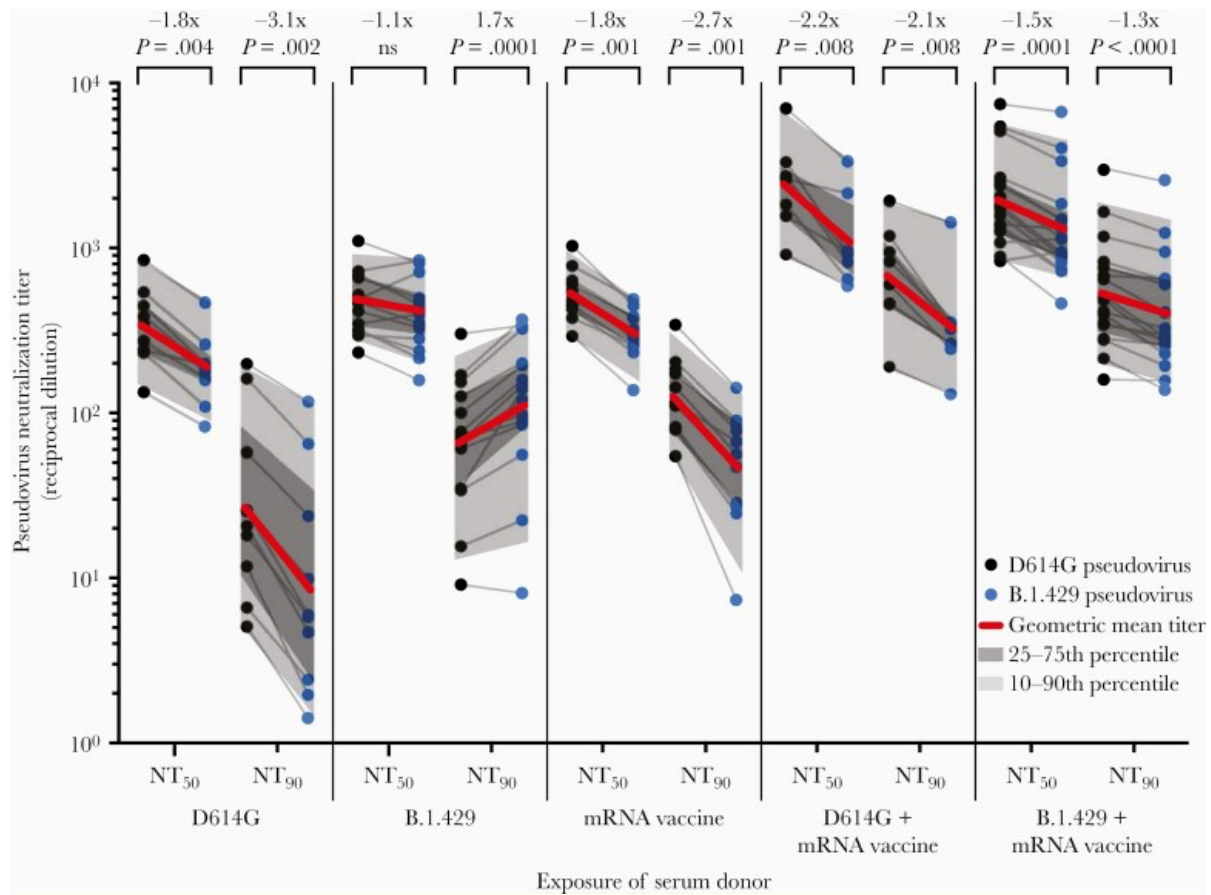
Figure 5.1. Neutralization of D614G and B.1.429 pseudoviruses by serum from individuals with different exposures. Plot of 50% and 90% pseudovirus neutralization titers (NT_{50} and NT_{90}) of serum samples obtained from donors with the indicated infection and/or vaccination exposures. Grey lines connect neutralization titer values for D614G (black dots) and B.1.429 (blue dots) pseudoviruses within each individual serum sample. Geometric mean neutralization titers for each serum group are marked with red lines and fold-change in NT_{50} and NT_{90} between D614G and B.1.429 pseudoviruses is shown along with P value. Dark grey shading marks the interquartile range of titer values in each group and light grey shading marks the 10th–90th percentile of the range. P values were calculated with a Wilcoxon matched-pairs signed-rank test.
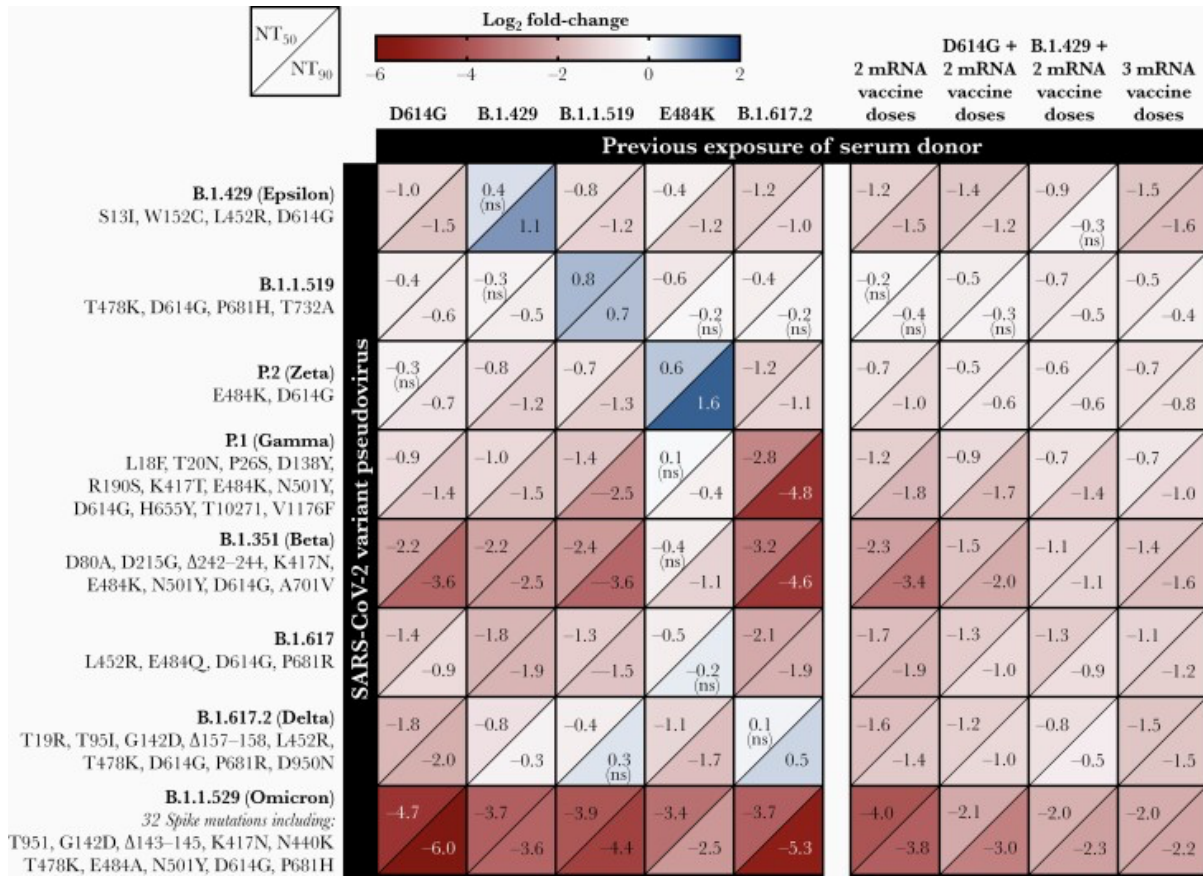
Figure 5.2. Change in variant pseudovirus neutralization titer relative to D614G. Matrix of normalized neutralization titers for 8 different variant pseudoviruses (rows) neutralized by 9 different pools of individual sera grouped by exposure (columns). Data are represented as a heat map of the log2 fold-change in NT_{50} (top left of each box) and NT_{90} (bottom right of each box) of each variant relative to D614G pseudovirus. All serum samples were collected at least 14 days after the date of the subject's positive COVID-19 test or date of most recent vaccine dose. All titer measurements are the mean of at least 3 independent experiments, each performed with 2 technical replicates. Positive log2 fold-change (blue) indicates an increase in neutralization titer for that variant relative to D614G pseudovirus, while negative log2 fold-change (red) indicates a decrease relative to D614G. Statistical significance was determined with unpaired t tests. All values are statistically significant (P value < .05) except where noted with ns to indicate the difference in variant neutralization titer is not significantly different from D614G pseudovirus neutralization titer in that serum pool. Abbreviations: COVID-19, coronavirus disease 2019; NT_{50} and NT_{90}, 50% and 90% neutralization titer.

Chapter 5 References:

1. Centers for Disease Control and Prevention. SARS-CoV-2 variant classifications and definitions. https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html. Accessed 15 December 2021.

2. Teyssou E, Delagrèverie H, Visseaux B, et al. The delta SARS-CoV-2 variant has a higher viral load than the beta and the historical variants in nasopharyngeal samples from newly diagnosed COVID-19 patients. *J Infect* 2021; 83:e1–3.

3. Greaney AJ, Starr TN, Barnes CO, et al. Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. *Nat Commun* 2021; 12:4196.

4. Liu J, Liu Y, Xia H, et al. BNT162b2-elicited neutralization of B.1.617 and other SARS-CoV-2 variants. *Nature* 2021; 596:273–5.

5. Corbett KS, Nason MC, Flach B, et al. Immune correlates of protection by mRNA-1273 vaccine against SARS-CoV-2 in nonhuman primates. Science 2021; 373:eabj0299.

6. Liu C, Ginn HM, Dejnirattisai W, et al. Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum. *Cell* 2021; 184:4220–36.e13.

7. Peng J, Liu J, Mann SA, et al. Estimation of secondary household attack rates for emergent spike L452R SARS-CoV-2 variants detected by genomic surveillance at a community-based testing site in San Francisco [published online ahead of print 31 March 2021]. *Clin Infect Dis* doi: 10.1093/cid/ciab283.

8. Kalantar KL, Carvalho T, de Bourcy CFA, et al. IDseq—an open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. *GigaScience* 2020; 9:giaa111.

9. Elledge SK, Zhou XX, Byrnes JR, et al. Engineering luminescent biosensors for point-of-care SARS-CoV-2 antibody detection. *Nat Biotechnol* 2021; 39:928–35.

10. Hoffmann M, Kleine-Weber H, Pöhlmann S. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol Cell* 2020; 78:779–84.e5.

11. Mattiuzzo G, Bentley EM, Hassall M, et al. *Establishment of the WHO International Standard and Reference Panel for anti-SARS-CoV-2 antibody*. Geneva, Switzerland: World Health Organization, 2020.

12. Liu C, Zhou D, Nutalai R, et al. The antibody response to SARS-CoV-2 beta underscores the antigenic distance to other variants [published online ahead of print 27 November 2021]. *Cell Host Microbe* doi: 10.1016/j.chom.2021.11.013.

13. McCallum M, Walls AC, Sprouse KR, et al. Molecular basis of immune evasion by the delta and kappa SARS-CoV-2 variants. *Science* 2021; 374:1621–6.

14. Wu K, Choi A, Koch M, et al. Variant SARS-CoV-2 mRNA vaccines confer broad neutralization as primary or booster series in mice. *Vaccine* 2021; 39:7394–400.

15. Corbett KS, Gagne M, Wagner DA, et al. Protection against SARS-CoV-2 beta variant in mRNA-1273 vaccine-boosted nonhuman primates. *Science* 2021; 374:1343–53.

Chapter 6: Functional investigation of SARS-CoV-2 host factors using Perturb-seq

This chapter is adapted from the bioRxiv preprint:

Sunshine S, Puschnik AS, Replogle JM, Laurie MT, Liu J, Zha BZ, Nuñez JK, Byrum JR, McMorrow AH, Frieman MB, Winkler J, Qiu X, Rosenberg OS, Leonetti MD, Ye CJ, Weissman JW, DeRisi JL, Hein MY. Systematic functional interrogation of SARS-CoV-2 host factors using Perturb-seq. bioRxiv **2022**; :2022.07.15.500120.

Supplemental files are included with original preprint.

Abstract

Genomic and proteomic screens have identified numerous host factors of SARS-CoV-2, but efficient delineation of their molecular roles during infection remains a challenge. Here we use Perturb-seq, a single-cell CRISPR screening approach, to investigate how inactivation of host factors changes the course of SARS-CoV-2 infection and the host response in human lung epithelial cells. Our high-dimensional data resolve complex phenotypes such as shifts in the stages of infection and modulations of the interferon response. However, only a small percentage of host factors showed such phenotypes upon perturbation. We further identified the NF-κB inhibitor IκBα (NFKBIA), as well as the translation factors EIF4E2 and EIF4H as strong host dependency factors acting early in infection. Overall, our study provides massively parallel functional characterization of host factors of SARS-CoV-2 and quantitatively defines their roles both in virus-infected and bystander cells.

Introduction

The coronavirus disease 2019 (COVID-19) pandemic, caused by SARS-CoV-2, has claimed millions of lives and remains a global health burden. Despite the success of rapid vaccine developments, barriers including vaccine access and uptake, as well as breakthrough infections make it imperative to develop both effective antivirals, and therapies targeting an overactive host immune response. A detailed understanding of the host determinants of infection, and the host response throughout infection will broadly inform efforts to develop novel antiviral agents.

Many studies have identified candidate host factors by an array of high-throughput methods, including protein-protein and protein-RNA interaction mapping, as well as CRISPR-based genetic screening [1–10]. Additionally, the host response to SARS-CoV-2 infection has been

71

investigated in single-cell transcriptional studies of blood, bronchial lavage, and tracheal aspirate from COVID-19 patients, human and animal (non-human primate, hamster, ferret) models of infection, and in cell lines infected in tissue culture [11–18]. However, it remains a challenge to validate individual candidate host factors, delineate their specific roles during infection, and evaluate their suitability as targets for interventions.

Here, we use Perturb-seq, a single-cell CRISPR screening approach [19–22], to understand how host genetic perturbations alter the course of SARS-CoV-2 infection and host transcriptional response in human lung epithelial cells. We designed a Perturb-seq library targeting 183 known coronavirus host factors. Those factors were either identified as host proteins interacting with viral factors, as protective hits in coronaviral genetic screens or for their known roles in antiviral host defense pathways [1,2,4–9,23–29]. We performed a CRISPR interference (CRISPRi) [30] screen in human lung carcinoma (Calu-3) cells, infected with a clinical isolate of SARS-CoV-2 collected in late 2020 (PANGO lineage B.1.503), and subsequently performed single-cell RNA sequencing, capturing both infected and uninfected bystander cells. Our results identify transcriptionally distinct clusters of infected and bystander cells, uncover new roles of genetic perturbations in interferon signaling, and functionally validate specific SARS-CoV-2 host dependency factors.

Results

Functional genomics of coronavirus host factors with a single-cell readout

To characterize the single-cell transcriptional response to SARS-CoV-2 infection and simultaneously test the effect of host genetic perturbations on viral RNA production and host response, we used Perturb-seq [19–22]. Perturb-seq combines CRISPR-based genetic perturbations with a rich, single-cell transcriptomics readout that is capable of capturing high-

dimensional phenotypes, making it well-suited for studying virus-host systems [31]. Viral infection leads to a heterogeneous response in a cell population, characterized, for instance, by cells being in different stages of infection and showing varying levels of activity of antiviral pathways [17,32,33]. Targeting critical host factors can cause shifts in the distribution of cellular states, which delivers insight into the function of any given host factor.

We performed our experiments in Calu-3 cells, a human respiratory epithelial cell line that endogenously expresses the entry receptor of SARS-CoV-2, ACE2, and has been previously used for several CRISPR screening and single-cell studies of SARS-CoV-2 [17,34]. We employed Calu-3 cells engineered to stably express the machinery for CRISPR interference (see Methods) [30,35,36]. CRISPRi is highly efficient at suppressing gene expression of selected targets without introducing double-strand breaks, with minimal off-target effects. On-target activity can be maximized by using two single-guide RNAs (sgRNAs) per target, expressed from one lentiviral vector [37,38].

We compiled a list of host factors from the literature on SARS-CoV-2 and other coronaviruses, mainly genes identified as protective hits in genetic screens for modifiers of SARS-CoV-2 or related coronavirus infections, and host proteins that were found to interact with viral proteins. We prioritized candidates with multiple lines of evidence supporting their roles in coronavirus biology. Additionally, we curated a list of factors involved in the innate immune response. Overall, we designed and cloned a library containing 239 elements, of which 195 target a single gene, 22 target combinations of two genes (typically paralogs or members of the same pathway, e.g. ACE2 + TMPRSS2 or IFNAR1 + IFNAR2), and 22 non-targeting controls (Table S1). We packaged the library into lentivirus and delivered it into the engineered Calu-3 cells at a low multiplicity of infection, followed by selection for cells with successful lentivirus integration.

73

We infected the resulting population for 24 h with a late-2020 clinical isolate of SARS-CoV-2 featuring only a single spike mutation, D614G, and 10 non-synonymous mutations in other genes relative to the ancestral isolate (PANGO lineage B.1.503, complete genome available at GISAID accession ID: EPI_ISL_13689582). Single-cell transcriptomes were then captured using a droplet-based microfluidic workflow (10x Genomics) with direct capture of sgRNAs to reveal which gene or gene pair was targeted in each cell [37] (Fig. 6.1A). After quality control filtering (Methods), we profiled the transcriptomes of 27,882 single cells which had exactly one Perturb-seq library element unambiguously assigned to them.

Transcriptional heterogeneity in SARS-CoV-2 infected cells

As a baseline for our subsequent Perturb-seq analysis, we first profiled the transcriptional response in the cell population upon infection, characterizing the spectrum of cellular states irrespective of the genetic perturbations present in the population. The heterogeneity of cellular states was primarily driven by the fraction of viral transcripts (Fig. 6.1B), which reached levels of up to 95% in some cells.

In order to compare infected and uninfected cells, we developed a classifier that determines the infection state of each cell based on the read counts of individual viral transcripts (Methods, Fig. 6.1C-D). Due to the presence of 'ambient' viral RNA, almost all cells have nonzero viral reads. To separate cells with true infection from those with spurious reads, the baseline of ambient viral RNA per cell was determined based on a spike-in of uninfected wild-type cells, which were identified by the absence of lentivirus-derived transcripts.

We sought to design an experimental strategy that captures single-cell transcriptomes of SARS-CoV-2 infected cells in a way that resolves both host and viral transcripts. Coronaviruses

have a unique transcript architecture [39], consisting of the (+)strand viral genome, numerous subgenomic mRNAs (sgmRNAs), and matching (-)sense counterparts. Importantly, all (+)sense transcripts start with the same ~72nt leader sequence at the 5′ end, followed by a junction to the body of the sgmRNAs. All (+)sense transcripts also share the same 3′ end and are polyadenylated. We reasoned that 3′ sequencing would not be able to resolve individual viral transcripts and therefore used the 10x Genomics 5′ workflow with a modified sequencing strategy that extends read1 to sequence from the 5′ end into the transcript, spanning the leader-body junction (Fig. 6.6A). A recent report found the same conceptual approach to maximize unambiguous detection of the different viral sgmRNAs [40].

Utilizing this 5′ sequencing strategy (see Methods), we resolved individual viral sgmRNAs and observed distinct patterns of viral transcript abundances in infected cells (Fig. 6.6B). The 3′-proximal Nucleocapsid (N) transcript was by far the most abundant viral RNA. Cell-by-cell correlation of the abundances of individual viral sgmRNAs was largely a function of genomic location: the abundances of the sgmRNAs proximal to N, encoding ORF3A, E, M, ORF6, ORF7ab, ORF8 showed the highest correlation with N. Conversely, the abundances of Spike and ORF1ab (i.e. whole genome) were much less correlated on a cell-by-cell basis. Additionally, we mapped the positions of leader-body junctions in sgmRNAs from our extended read1 data and found both the positions as well as their relative frequencies of individual junctions to be in agreement with measurements derived from bulk, whole-transcript sequencing data [39] (Fig. 6.6C).

Next, using our infection state classification, we observed the cell cycle phase as a major contributor to the heterogeneity among uninfected cells, with subclusters often representing cells within one predominant phase. Conversely, infected cells showed a pronounced, general shift in

their cell cycle phases: we observe far fewer infected cells in S phase and the proportion of G1 cells increased approximately two-fold (Fig. 6.1E-F), suggesting that cell cycle arrest occurs upon infection. Furthermore, infected and uninfected bystander cells differed dramatically in the total amount of detectable cellular RNA, quantified by the number of unique molecular identifiers (UMIs) per cell, indicating a pronounced shutoff of host gene expression in infected cells (Fig. 6.1G). This observation is consistent with a recent study showing that SARS-CoV-2 NSP1 specifically degrades transcripts lacking the viral 5′ leader sequence, enabling the virus to dominate the cellular mRNA pool [41].

To further characterize the heterogeneity within the infected and bystander populations, different cell states were delineated using Leiden clustering, defining 12 clusters of bystander cells (clusters A–L) and 7 clusters of infected cells (clusters M–S) (Fig. 6.2A).

To identify transcriptional patterns within these different clusters, we evaluated gene expression within each cluster. (Fig. 6.2B). Bystander cells (clusters A–L) varied in their expression of genes associated with antigen presentation, chemokines, and interferon-stimulated genes (ISGs). ISGs including IFI6, IFI27, and ISG15 (Fig. 6.7A) were prominently more abundant in bystander cells compared to infected cells. This suggests active suppression of the interferon response in infected cells, a phenomenon that has been observed for many different viruses [31,33,41,42].

We identified a small but prominent subset of cells (bystander cluster L and infected cluster M) expressing interferon β (IFNB1) and λ (IFNL1/2/3) (Fig. 6.7B) and a number of chemokines (CXCL1/2/3/10/11, CCL5, IL6, CXCL8/IL8). This observation is consistent with prior single-cell work showing a subset of interferon-producing cells after SARS-CoV-2 infection [17], and studies that assessed interferon production in bulk [43]. Notably, all interferon-producing cells exhibited

pronounced expression of both NF-κB pathway genes and ISGs. Additionally, this population expressed genes associated with antigen presentation and translation regulation/stress response (e.g. PPP1R15A). These features were reminiscent of subpopulations of abortively infected cells which have been characterized for the herpesviruses HSV-1 and HCMV [31,33]. However, only approximately 20% of the interferon-producing cells in our dataset were classified as infected based on the abundances of viral transcripts (cluster M).

Infected cells (clusters M–S) varied not only in their fractions of viral transcripts, but also showed a concomitant shift in cell cycle distribution (Fig. 6.2A), and subtle host transcriptional patterns (Fig. 6.2B). A number of host transcripts were generally upregulated in infected cells, including genes associated with NF-κB signaling such as NFKBIA (Fig. 6.7C), NFKBIE/Z, EGR1, REL and RELB (Fig. 6.2E). In addition, genes related to cell stress (ATF3, FOS, JUN) were upregulated in most infected clusters.

It is conceivable that the apparent downregulation of some transcripts in infected cells (such as ISGs) is an artifact caused by the global host shutoff. Therefore, we repeated gene expression and cell cycle analyses on cells that were downsampled to the read depth of infected cells (bottom 2% of the UMI distribution). These data recapitulate our prior findings and suggest that despite host shutoff, we were able to detect transcriptional changes in infected and bystander cells (Fig. 6.7D-I).

Host perturbations alter infection dynamics

To determine how the activity of host factors affects the response of a cell population to SARS-CoV-2 infection, we next evaluated how each genetic perturbation in our CRISPRi library altered viral load and bystander activation. To ensure sufficient representation of our 239 library

elements, we assessed the distribution of captured cells for these elements and determined the peak

of that distribution (mode) to be at 138 cells (Fig. 6.8A). 48 library elements had less than 55 cells

each, forming a distinct lower mode in the distribution of cell numbers, suggesting that they target

genes essential for the growth of Calu-3 cells. As these elements lacked appropriate coverage for

proper evaluation of infection dynamics, they were removed, resulting in 25,835 remaining cells,

on which we based all downstream analyses (Table S2). Among well-represented targets, the

median knockdown efficiency was 91%, and 90% of our library showed greater than 75%

knockdown of their respective target transcripts in uninfected cells confirming the efficacy of

CRISPRi targeting in Calu-3 cells (Fig. 6.8B).

To test which host factors confer protection from infection upon perturbation, we compared

the distributions of viral loads in cells with any given CRISPRi target against the population of

cells with non-targeting controls (Fig. 6.3A). Knockdown of only one factor, SEC62, resulted in

increased viral loads. This was unexpected in light of genetic screens that identified SEC62

knockout as protective against human coronavirus (HCoV) OC43 infection [9]. SEC62 is involved

in the post-translational targeting of proteins across the endoplasmic reticulum, acts as an

autophagy receptor in the ER, and is a known interactor of SEC61B [44,45].. On the other hand,

knockdown of the known entry factors ACE2 and TMPRSS2, both alone and in combination, led

to strongly reduced viral loads. Similarly, TMPRSS2 in combination with either Furin, Cathepsin

B, or L (but notably not Furin or either Cathepsin alone) resulted in substantially reduced fractions

of viral RNA, suggesting partial redundancy of those entry factors. Knockdown of BRD2 also

reduced viral loads considerably, which is consistent with the recent finding that BRD2 is required

for efficient transcription of ACE2 [35].

Aside from those known factors involved in viral entry, we identified a number of additional, strongly protective factors such as: the autophagy factor ATG14 [46], as well as translation factors EIF4E2 (4EHP) and EIF4H. Translation factors EIF4E2 and EIF4H were previously found to interact with the viral proteins NSP2 and NSP9, respectively [1,2]. EIF4E2 represses translation initiation by binding to the mRNA cap and can be ISGylated to enhance this cap-binding activity [47]. In the setting of SARS-CoV-2, EIF4E2 surfaced as an unvalidated protective host factor in one genetic screen [5]. The second translation factor that conferred protection from infection upon knockdown, EIF4H, binds to and stimulates RNA helicase activity of EIF4A [48,49]. Additionally, EIF4H is reported to interact with SARS-CoV-2 RNA [1,5,50]. Notably, the protective phenotypes when targeting EIF4E2 and EIF4H do not appear to reflect a general effect of perturbing translation factors, as EIF4B did not significantly alter infection dynamics.

Additionally, knockdown of VMP1 and MPP5 conferred protection from infection. VMP1 is involved in cytoplasmic vacuole formation and autophagosome assembly, when it interacts with TMEM41B [51], a known pan-coronavirus host dependency factor [5,9]. MPP5 is involved in tight junction formation and was similarly identified as a protective hit in a genome-wide CRISPR survival screen [9], and as interactor of the E protein of SARS-CoV-1 [52]. Our data validate those proteins as protective host factors.

Lastly, we observed that knockdown of the NF-κB inhibitor IκBα (encoded by NFKBIA) significantly reduced viral loads. The NF-κB pathway is well-known to be activated in the setting of viral infections [53], and its activity was reported to be important for SARS-CoV-2 replication [54]. While NFKBIA is transcriptionally upregulated in SARS-CoV-2 infected cells as shown in

our data (Fig. 6.7C) and by others [17,54,55], it has not appeared in any genetic screen to our knowledge as a protective factor.

At baseline, IκBα inhibits the NF-κB pathway by binding to and retaining p65/RELA-containing complexes in the cytosol [56]. Canonical pathway activation induces proteasomal degradation of IκBα/NFKBIA, leading to p65/RELA nuclear translocation and subsequent transcription of NF-κB target genes (including NFKBIA, forming a negative feedback loop). Prior studies have shown that the papain-like proteases (PLPro) of both SARS-CoV-1 and SARS-CoV-2 can deubiquitylate and thereby stabilize IκBα, thus decreasing p65/RELA nuclear translocation and suppressing NF-κB pathway activation [55,57].

However, our data show that knockdown of NFKBIA does not lead to transcriptional activation of the NF-κB pathway in bystander cells (Fig. 6.9A), arguing against constitutive activation as a phenotypic outcome. Knocking down RELA or RELB, both individually or in combination, did not result in a protective phenotype. Our data suggest a dependency of SARS-CoV-2 on NFKBIA, which may be independent of its inhibitory role in the NF-κB pathway.

To further investigate the phenotypic response of NFKBIA perturbation, we utilized the OpenCell collection of HEK293T cell lines expressing split mNeonGreen (mNG)-tagged proteins from their endogenous loci [45]. First, we confirmed that the NF-κB pathway was functional in cells expressing mNG-tagged RELA. Using live-cell fluorescent microscopy, we observed the expected p65/RELA translocation to the nucleus after TNF-α stimulation (Fig. 6.9B). We then generated polyclonal NFKBIA knockout lines in the background of the mNG-RELA line. Without stimulation, there was no constitutive p65/RELA translocation to the nucleus in NFKBIA KO cells. After treatment with TNF-α, we observed a blunted response with delayed and incomplete p65/RELA nuclear translocation in NFKBIA KO cells compared to control cells (Fig. 6.9B). These

data are in agreement with prior studies that show a delayed response to NF-κB pathway stimulation in the setting of an NFKBIA knockout [58], and suggest a compensatory mechanism that prevents both constitutive and acute pathway activation.

Next, we orthogonally validated the observed protective phenotypes of inactivated NFKBIA, EIF4E2 and EIF4H by generating knockout lines from Huh-7.5.1 (hepatocellular carcinoma) cells ectopically expressing ACE2 and TMPRSS2 (Fig. 6.3B). This cell line is permissive for SARS-CoV-2 infection and has been used for pooled CRISPR screening [9]. We infected polyclonal pools of knockout cells with SARS-CoV-2 and quantified the fraction of infected cells by fluorescence microscopy, staining for the viral nucleocapsid protein. Compared to non-targeting controls, ACE2, NFKBIA, EIF4E2, and EIF4H knockout cell lines showed a substantial decrease in infection (Fig 3C-D). NFKBIA knockout cells displayed a 31.8% decrease in infection, 85.5% in EIF4E2 KO, and 33.2% in EIF4H KO cells compared to non-targeting control cells. This confirms the findings from our Perturb-seq data and implicates that all those factors play a role in SARS-CoV-2 infection in different cell types.

Systematic classification of host factor phenotypes

Changes in the viral load distribution is only one manifestation of the multitude of cellular phenotypes resulting from host factor perturbation. To achieve a systematic and unbiased characterization of host factor perturbation phenotypes, beyond viral protection/sensitization, we monitored how different perturbations shift the proportion of cells in distinct cellular states. Qualitatively, this can be assessed by looking the distribution of cells with a given genetic perturbation on the UMAP projection. More quantitatively, one can count cells in the different Leiden clusters and determine how a given host factor perturbation changes the relative numbers

of cells by cluster (see Fig. 6.2A). This approach not only identifies host factor perturbations that alter cellular states and sorts them by similarity, but also narrows down the underlying mechanism by directly pinpointing the cellular states that are affected by the perturbation [31].

Cells with non-targeting control sgRNAs were uniformly distributed across the UMAP representation of the cell population (Fig. 6.4A). In comparison, cells with certain genetic perturbations deviated from this pattern in specific ways. First, cells with sgRNAs targeting known entry factors were specifically excluded from all infected clusters in UMAP space (Fig. 6.4A). The same was also true for cells with NFKBIA-targeting sgRNAs. Moreover, those cells were similarly excluded from two clusters (bystander clusters F and L, the latter being one of the interferon-producing clusters). These clusters border infected clusters and were classified as uninfected based on viral transcripts in quantities below noise level. Based on the observation that entry factor inactivation excludes cells from these clusters, we speculate that these two clusters represent cells that are in the earliest stage of infection, have been infected with a defective viral particle, or are in a state where transcription of viral genes is effectively suppressed by an antiviral host response.

We systematically quantified the under/overrepresentation of cells with a given host factor perturbation in individual clusters (Fig. 6.4B), compared to cells with non-targeting control sgRNAs. The results can be visualized as a heatmap of the odds-ratio of how targeting a certain host factor changes the occupancy of each cluster (Fig. 6.4C, Fig. 6.10A), which can further be projected onto a UMAP of host factor phenotypes (Fig. 6.4D, Fig. 6.10B).

This analysis re-confirmed the group of proviral factors, which are strongly protective when inactivated (Fig. 6.4D, 6.10B, blue highlight). A second group of perturbations which caused a distinct re-distribution of cells across the individual clusters were cells with inactivated members of the interferon pathway (Fig. 6.4A, 4D, 6.10B, orange highlight). Those cells were shifted from

82

bystander clusters representing cells with high expression of ISGs (clusters B, D, E, F) to the cluster with a low degree of interferon response (cluster A) (see Fig. 6.10A, 2B).

The group of interferon signaling factors contained not only expected genes (IFNAR2, STAT2, IRF3, IRF9), but also genes not routinely implicated in the interferon response such as SPNS1, KEAP1 and GPR89A/B. To evaluate these in more detail, we scored the extent of interferon response in single cells based on a previously established list of ISGs that are readily detected by single cell RNA sequencing [31]. We subsequently tested for statistically significant shifts in this interferon score for each perturbation compared to non-targeting controls (Fig. 6.5A, B). To rule out the effect of viral antagonism of this pathway, we limited this analysis to bystander cells.

Knockdown of GPR89A/B, KEAP1, SPNS1 and BRD2 significantly decreased bystander activation as measured by our ISG scores, confirming these proteins as regulators of the interferon signaling pathway. GPR89A and GPR89B are sequence-identical paralogs, encoding a G protein coupled receptor, and proteomic studies report interactions of this protein with multiple SARS-CoV-2 proteins (M, NSP6 and ORF7B) [4]. Notably, GPR89A/B overexpression is reported to activate the NF-κB signaling pathway [59], and this protein is thought to be important for Golgi acidification and glycosylation [60]. KEAP1 is a repressor of NRF2, which acts as a regulator of the inflammatory response [61]. Our findings for KEAP1 are consistent with prior work that showed repression of inflammatory genes in *Keap1* deficient murine cells [62]. SPNS1 is involved in lipid and transmembrane transport, and Wang et al. reported that genetic knockout of this gene protects from hCoV-229E and hCoV-OC43 infections *in vitro* [9]. Both KEAP1 and SPNS1 were shown to interact with SARS-CoV-2 ORF3 and ORF7b, respectively [1,4]. Furthermore, CRISPRi knockdown of BRD2 decreased the overall sensitivity of bystander cells in our study, which aligns

with prior reports that perturbation of BRD2 reduces interferon signaling [35]. Taken together, our analytical framework identified genes not routinely implicated in bystander activation, and proved to be very sensitive to identify factors with subtle phenotypes beyond strong protection from infection.

Discussion

In this study, we measured the dynamics of SARS-CoV-2 infection in tissue culture, and simultaneously validated and functionally characterized host factors of infection. Perturb-seq delivers a high-dimensional, phenotypic single-cell readout, characterizing both the intrinsic heterogeneity of a SARS-CoV-2 infected population, and the response to many host factor perturbations. We captured different functional outcomes and simultaneously classified host factors by the similarities of their roles during infection and bystander activation. Our study thereby complements and greatly expands upon the genomic and proteomic screens which initially informed our selection of host factors included in our Perturb-seq library [1,2,4–6].

Our transcriptional analysis revealed upregulation of key NF-κB pathway members, including NFKBIA, in SARS-CoV-2 infected cells. This is consistent with findings of earlier studies [17,18,54]. Considering the prominent transcriptional host shutoff, we speculate that viral factors trigger the upregulation of NFKBIA and/or protect the transcript from degradation. Moreover, our study demonstrates that IκBα/NFKBIA can be targeted genetically to confer strong protection from SARS-CoV-2 infection. Our data suggest blunting of NF-κB pathway activation as one underlying mechanism to explain this phenotype. While somewhat counterintuitive in light of IκBα/NFKBIA's role as an NF-κB inhibitor, this result is in line with data from previous high-throughput optical imaging screens [58]. Furthermore, we suspect that independent from IκBα's

canonical inhibitory role in NF-κB signaling, IκBα may additionally be co-opted in another way that benefits viral proliferation. This is corroborated by a recent report that overexpression of a dominant-negative IκBα mutant enhances SARS-CoV-2 infection in A549 cells, while simultaneously reducing p65/RELA nuclear translocation [82].

Our data further establishes that two translation factors, EIF4E2 and EIF4H, are required for SARS-CoV-2 infection. While prior studies report that both factors interact with viral proteins [1], here, we show that knockdown and knockout of these factors decreases infection. The 4EHP(EIF4E2)-GIGYF2 complex is involved in ribosome-associated quality control by preventing translation initiation of faulty mRNA [63–65], and its interaction with NSP2 is conserved across SARS-CoV-1, SARS-CoV-2 and MERS-CoV [2]. Others have proposed that viral NSP2 interacts with the 4EHP(EIF4E2)-GIGYF2 complex to inhibit host translation initiation [66]. However, the strongly protective knockdown phenotype of EIF4E2 observed in our data leads us to instead hypothesize that binding of viral NSP2 to EIF4E2 drives preferential translation of viral RNA. In this manner, the virus may subvert what is normally a defense mechanism for its exclusive use within the cell. Further investigation to determine which transcripts EIF4E2 binds to in the setting of infection will aid our understanding of the underlying mechanism of EIF4E2 utilization by coronaviruses.

EIF4H directly binds to and stimulates the DEAD box RNA helicase EIF4A [49]. A pharmacological inhibitor of EIF4A, Zotatafin, decreases SARS-CoV-2 infection *in vitro*, and clinical trials (NCT04632381) are underway to evaluate its safety and efficacy in humans [1,67]. Our experiments reveal a viral dependency on the EIF4A binding partner EIF4H, suggesting a complementary, and possibly synergistic point for additional therapeutic intervention.

Finally, our systematic characterization of each genetic perturbation revealed regulators of bystander activation. KEAP1, GPR89A/B, and SPNS1, which were previously found to be protective when knocked out [5,7,9], did not alter infection dynamics within our study. We speculate that knockout of these genes was identified as protective in survival screens due to their lack of interferon sensitivity, leading to protection from interferon-induced death [5,9]. Conversely, it is possible that these contrary phenotypes are representative of the different timeframes of our Perturb-seq experiments (24 hours) compared to genetic survival screens (7+ days). While knockdown of KEAP1, GPR89A/B, and SPNS1 initially decreased interferon stimulation in our experiments, it is conceivable that these factors have a secondary role in protecting the population from infection in long-term cultures.

While our Perturb-seq library was designed to include genes with experimental evidence of roles in coronavirus biology, only ~13% of these factors ultimately showed significant phenotypes during the first 24 hours of infection in our cell culture model. This underscores the necessity for high-throughput orthogonal validation and characterization of host factors in different cell types. We do expect that specific host factor perturbation phenotypes, in particular of factors acting at the later stages of the viral life cycle such as virion assembly and egress, cannot be resolved by Perturb-seq. Similarly, host factors that are active only in rare subsets of cells, such as the interferon-producing subpopulation, may be difficult for Perturb-seq to dissect without increasing the scale of these experiments.

In summary, our study presents comprehensive transcriptional profiling of SARS-CoV-2 infection dynamics, tests the effect of 183 host factor perturbations on infection, and characterizes the host response of each perturbation. Key advances of this work include the identification of genes involved in bystander activation and functional validation of host dependencies factors of

SARS-CoV-2. Our study highlights the utility of Perturb-seq for large-scale systematic characterization of host factors essential for pathogen infections and establishes the groundwork for future mechanistic studies to investigate how SARS-CoV-2 modulates both the NF-κB pathway and translation.

Chapter 6 References

1. Gordon DE, Jang GM, Bouhaddou M, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature. **2020**; 583(7816):459–468.

2. Gordon DE, Hiatt J, Bouhaddou M, et al. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. Science [Internet]. **2020**; 370(6521). Available from: http://dx.doi.org/10.1126/science.abe9403

3. Schmidt N, Lareau CA, Keshishian H, et al. The SARS-CoV-2 RNA–protein interactome in infected human cells [Internet]. Nature Microbiology. 2021. p. 339–353. Available from: http://dx.doi.org/10.1038/s41564-020-00846-z

4. Stukalov A, Girault V, Grass V, et al. Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. Nature. **2021**; 594(7862):246–252.

5. Schneider WM, Luna JM, Hoffmann H-H, et al. Genome-Scale Identification of SARS-CoV-2 and Pan-coronavirus Host Factor Networks. Cell. **2021**; 184(1):120–132.e14.

6. Hoffmann H-H, Sánchez-Rivera FJ, Schneider WM, et al. Functional interrogation of a SARS-CoV-2 host protein interactome identifies unique and shared coronavirus host factors. Cell Host Microbe. **2021**; 29(2):267–280.e5.

7. Wei J, Alfajaro MM, DeWeirdt PC, et al. Genome-wide CRISPR Screens Reveal Host Factors Critical for SARS-CoV-2 Infection. Cell. **2021**; 184(1):76–91.e13.

8. Daniloski Z, Jordan TX, Wessels H-H, et al. Identification of Required Host Factors for SARS-CoV-2 Infection in Human Cells. Cell. **2021**; 184(1):92–105.e16.

9. Wang R, Simoneau CR, Kulsuptrakul J, et al. Genetic Screens Identify Host Factors for SARS-CoV-2 and Common Cold Coronaviruses. Cell. **2021**; 184(1):106–119.e14.

10. Flynn RA, Belk JA, Qi Y, et al. Discovery and functional interrogation of SARS-CoV-2

RNA-host protein interactions. Cell. **2021**; 184(9):2394–2411.e16.

11. Wijst MGP van der, Vazquez SE, Hartoularos GC, et al. Type I interferon autoantibodies are associated with systemic immune alterations in patients with COVID-19. Sci Transl Med. **2021**; 13(612):eabh2624.

12. Liao M, Liu Y, Yuan J, et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. Nat Med. **2020**; 26(6):842–844.

13. Sarma A, Christenson SA, Byrne A, et al. Tracheal aspirate RNA sequencing identifies distinct immunological features of COVID-19 ARDS. Nat Commun. **2021**; 12(1):5152.

14. Speranza E, Williamson BN, Feldmann F, et al. Single-cell RNA sequencing reveals SARS-CoV-2 infection dynamics in lungs of African green monkeys. Sci Transl Med [Internet]. **2021**; 13(578). Available from: http://dx.doi.org/10.1126/scitranslmed.abe8146

15. Nouailles G, Wyler E, Pennitz P, et al. Temporal omics analysis in Syrian hamsters unravel cellular effector responses to moderate COVID-19. Nat Commun. **2021**; 12(1):4869.

16. Lee JS, Koh J-Y, Yi K, et al. Single-cell transcriptome of bronchoalveolar lavage fluid reveals sequential change of macrophages during SARS-CoV-2 infection in ferrets. Nat Commun. **2021**; 12(1):4567.

17. Wyler E, Mösbauer K, Franke V, et al. Transcriptomic profiling of SARS-CoV-2 infected human cell lines identifies HSP90 as target for COVID-19 therapy. iScience. **2021**; 24(3):102151.

18. Ravindra NG, Alfajaro MM, Gasque V, et al. Single-cell longitudinal analysis of SARS-CoV-2 infection in human airway epithelium identifies target cells, alterations in gene expression, and cell state changes. PLoS Biol. **2021**; 19(3):e3001143.

19. Adamson B, Norman TM, Jost M, et al. A Multiplexed Single-Cell CRISPR Screening

Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell. **2016**; 167(7):1867–1882.e21.

20. Dixit A, Parnas O, Li B, et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell. **2016**; 167(7):1853–1866.e17.

21. Datlinger P, Rendeiro AF, Schmidl C, et al. Pooled CRISPR screening with single-cell transcriptome readout. Nat Methods. **2017**; 14(3):297–301.

22. Jaitin DA, Weiner A, Yofe I, et al. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. Cell. **2016**; 167(7):1883–1896.e15.

23. Lim YX, Ng YL, Tam JP, Liu DX. Human Coronaviruses: A Review of Virus-Host Interactions. Diseases [Internet]. **2016**; 4(3). Available from: http://dx.doi.org/10.3390/diseases4030026

24. Wong HH, Kumar P, Tay FPL, Moreau D, Liu DX, Bard F. Genome-Wide Screen Reveals Valosin-Containing Protein Requirement for Coronavirus Exit from Endosomes. J Virol. **2015**; 89(21):11116–11128.

25. V'kovski P, Gerber M, Kelly J, et al. Determination of host proteins composing the microenvironment of coronavirus replicase complexes by proximity-labeling. Elife [Internet]. **2019**; 8. Available from: http://dx.doi.org/10.7554/eLife.42037

26. Wilde AH de, Wannee KF, Scholte FEM, et al. A Kinome-Wide Small Interfering RNA Screen Identifies Proviral and Antiviral Host Factors in Severe Acute Respiratory Syndrome Coronavirus Replication, Including Double-Stranded RNA-Activated Protein Kinase and Early Secretory Pathway Proteins. J Virol. **2015**; 89(16):8318–8333.

27. Pfefferle S, Schöpf J, Kögl M, et al. The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors. PLoS Pathog. **2011**;

7(10):e1002331.

28. Biering SB, Sarnik SA, Wang E, et al. Genome-wide bidirectional CRISPR screens identify mucins as host factors modulating SARS-CoV-2 infection. Nat Genet [Internet]. **2022**; . Available from: http://dx.doi.org/10.1038/s41588-022-01131-x

29. Rebendenne A, Roy P, Bonaventure B, et al. Bidirectional genome-wide CRISPR screens reveal host factors regulating SARS-CoV-2, MERS-CoV and seasonal HCoVs. Nat Genet [Internet]. **2022**; . Available from: http://dx.doi.org/10.1038/s41588-022-01110-2

30. Gilbert LA, Horlbeck MA, Adamson B, et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell. **2014**; 159(3):647–661.

31. Hein MY, Weissman JS. Functional single-cell genomics of human cytomegalovirus infection. Nat Biotechnol. **2022**; 40(3):391–401.

32. Zanini F, Robinson ML, Croote D, et al. Virus-inclusive single-cell RNA sequencing reveals the molecular signature of progression to severe dengue. Proc Natl Acad Sci U S A. **2018**; 115(52):E12363–E12369.

33. Drayman N, Patel P, Vistain L, Tay S. HSV-1 single-cell analysis reveals the activation of anti-viral and developmental programs in distinct sub-populations. Elife [Internet]. **2019**; 8. Available from: http://dx.doi.org/10.7554/eLife.46339

34. Biering SB, Sarnik SA, Wang E, et al. Genome-wide, bidirectional CRISPR screens identify mucins as critical host factors modulating SARS-CoV-2 infection [Internet]. bioRxiv; 2021. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.04.22.440848

35. Samelson AJ, Tran QD, Robinot R, et al. BRD2 inhibition blocks SARS-CoV-2 infection by reducing transcription of the host cell receptor ACE2. Nat Cell Biol. **2022**; 24(1):24–34.

36. Qi LS, Larson MH, Gilbert LA, et al. Repurposing CRISPR as an RNA-guided platform for

sequence-specific control of gene expression. Cell. **2013**; 184(3):844.

37. Replogle JM, Norman TM, Xu A, et al. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. Nat Biotechnol. **2020**; 38(8):954–961.

38. Replogle JM, Saunders RA, Pogson AN, et al. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. Cell [Internet]. **2022**; . Available from: http://dx.doi.org/10.1016/j.cell.2022.05.013

39. Kim D, Lee J-Y, Yang J-S, Kim JW, Kim VN, Chang H. The Architecture of SARS-CoV-2 Transcriptome. Cell. **2020**; 181(4):914–921.e10.

40. Cohen P, DeGrace EJ, Danziger O, Patel RS, Rosenberg BR. Unambiguous detection of SARS-CoV-2 subgenomic mRNAs with single cell RNA sequencing. bioRxiv [Internet]. **2021**; . Available from: http://dx.doi.org/10.1101/2021.11.22.469642

41. Finkel Y, Gluck A, Nachshon A, et al. SARS-CoV-2 uses a multipronged strategy to impede host protein synthesis. Nature. **2021**; 594(7862):240–245.

42. Xia H, Cao Z, Xie X, et al. Evasion of Type I Interferon by SARS-CoV-2. Cell Rep. **2020**; 33(1):108234.

43. Yin X, Riva L, Pu Y, et al. MDA5 Governs the Innate Immune Response to SARS-CoV-2 in Lung Epithelial Cells. Cell Rep. **2021**; 34(2):108628.

44. Fumagalli F, Noack J, Bergmann TJ, et al. Translocon component Sec62 acts in endoplasmic reticulum turnover during stress recovery. Nat Cell Biol. **2016**; 18(11):1173–1184.

45. Cho NH, Cheveralls KC, Brunner A-D, et al. OpenCell: Endogenous tagging for the cartography of human cellular organization. Science. **2022**; 375(6585):eabi6983.

46. Shang C, Zhuang X, Zhang H, et al. Inhibition of Autophagy Suppresses SARS-CoV-2

Replication and Ameliorates Pneumonia in hACE2 Transgenic Mice and Xenografted Human Lung Tissues. J Virol. **2021**; 95(24):e0153721.

47. Okumura F, Zou W, Zhang D-E. ISG15 modification of the eIF4E cognate 4EHP enhances cap structure-binding activity of 4EHP. Genes Dev. **2007**; 21(3):255–260.

48. Rogers GW Jr, Richter NJ, Lima WF, Merrick WC. Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F. J Biol Chem. **2001**; 276(33):30914–30922.

49. Richter NJ, Rogers GW Jr, Hensold JO, Merrick WC. Further biochemical and kinetic characterization of human eukaryotic initiation factor 4H. J Biol Chem. **1999**; 274(50):35415–35424.

50. Labeau A, Fery-Simonian L, Lefevre-Utile A, et al. Characterization and functional interrogation of the SARS-CoV-2 RNA interactome [Internet]. Cell Reports. 2022. p. 110744. Available from: http://dx.doi.org/10.1016/j.celrep.2022.110744

51. Morita K, Hama Y, Izume T, et al. Genome-wide CRISPR screen identifies TMEM41B as a gene required for autophagosome formation [Internet]. Journal of Cell Biology. 2018. p. 3817–3828. Available from: http://dx.doi.org/10.1083/jcb.201804132

52. Teoh K-T, Siu Y-L, Chan W-L, et al. The SARS coronavirus E protein interacts with PALS1 and alters tight junction formation and epithelial morphogenesis. Mol Biol Cell. **2010**; 21(22):3838–3852.

53. Rahman MM, McFadden G. Modulation of NF-κB signalling by microbial pathogens. Nat Rev Microbiol. **2011**; 9(4):291–306.

54. Nilsson-Payant BE, Uhl S, Grimont A, et al. The NF-κB Transcriptional Footprint Is Essential for SARS-CoV-2 Replication. J Virol. **2021**; 95(23):e0125721.

55. Shin D, Mukherjee R, Grewe D, et al. Papain-like protease regulates SARS-CoV-2 viral

spread and innate immunity. Nature. **2020**; 587(7835):657–662.

56. Bonizzi G, Karin M. The two NF-kappaB activation pathways and their role in innate and adaptive immunity. Trends Immunol. **2004**; 25(6):280–288.

57. Frieman M, Ratia K, Johnston RE, Mesecar AD, Baric RS. Severe acute respiratory syndrome coronavirus papain-like protease ubiquitin-like domain and catalytic domain regulate antagonism of IRF3 and NF-kappaB signaling. J Virol. **2009**; 83(13):6689–6705.

58. Feldman D, Singh A, Schmid-Burgk JL, et al. Optical Pooled Screens in Human Cells. Cell. **2019**; 179(3):787–799.e17.

59. Matsuda A, Suzuki Y, Honda G, et al. Large-scale identification and characterization of human genes that activate NF-κB and MAPK signaling pathways [Internet]. Oncogene. 2003. p. 3307–3318. Available from: http://dx.doi.org/10.1038/sj.onc.1206406

60. Maeda Y, Ide T, Koike M, Uchiyama Y, Kinoshita T. GPHR is a novel anion channel critical for acidification and functions of the Golgi apparatus. Nat Cell Biol. **2008**; 10(10):1135–1145.

61. Thimmulappa RK, Lee H, Rangasamy T, et al. Nrf2 is a critical regulator of the innate immune response and survival during experimental sepsis. J Clin Invest. **2006**; 116(4):984–995.

62. Gunderstofte C, Iversen MB, Peri S, et al. Nrf2 Negatively Regulates Type I Interferon Responses and Increases Susceptibility to Herpes Genital Infection in Mice. Front Immunol. **2019**; 10:2101.

63. Hickey KL, Dickson K, Cogan JZ, et al. GIGYF2 and 4EHP Inhibit Translation Initiation of Defective Messenger RNAs to Assist Ribosome-Associated Quality Control. Mol Cell. **2020**; 79(6):950–962.e6.

64. Juszkiewicz S, Slodkowicz G, Lin Z, Freire-Pritchett P, Peak-Chew S-Y, Hegde RS. Ribosome collisions trigger cis-acting feedback inhibition of translation initiation. Elife [Internet]. **2020**; 9. Available from: http://dx.doi.org/10.7554/eLife.60038

65. Morita M, Ler LW, Fabian MR, et al. A novel 4EHP-GIGYF2 translational repressor complex is essential for mammalian development. Mol Cell Biol. **2012**; 32(17):3585–3593.

66. Gupta M, Azumaya CM, Moritz M, et al. CryoEM and AI reveal a structure of SARS-CoV-2 Nsp2, a multifunctional protein involved in key host processes. bioRxiv [Internet]. **2021**; . Available from: http://dx.doi.org/10.1101/2021.05.10.443524

67. White KM, Rosales R, Yildiz S, et al. Plitidepsin has potent preclinical efficacy against SARS-CoV-2 by targeting the host protein eEF1A. Science. **2021**; 371(6532):926–931.

68. Honko AN, Storm N, Bean DJ, Vasquez JH, Downs SN, Griffiths A. Rapid Quantification and Neutralization Assays for Novel Coronavirus SARS-CoV-2 Using Avicel RC-591 Semi-Solid Overlay [Internet]. Available from: http://dx.doi.org/10.20944/preprints202005.0264.v1

69. Rambaut A, Holmes EC, O'Toole Á, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. **2020**; 5(11):1403–1407.

70. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. **2018**; 19(1):15.

71. Cong L, Ran FA, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. **2013**; 339(6121):819–823.

72. Stirling DR, Swain-Bowden MJ, Lucas AM, Carpenter AE, Cimini BA, Goodman A. CellProfiler 4: improvements in speed, utility and usability. BMC Bioinformatics. **2021**; 22(1):433.

73. Edelstein AD, Tsuchida MA, Amodaj N, Pinkard H, Vale RD, Stuurman N. Advanced methods of microscope control using μManager software. J Biol Methods [Internet]. **2014**; 1(2). Available from: http://dx.doi.org/10.14440/jbm.2014.36

74. Sofroniew N, Lambert T, Evans K, et al. napari: a multi-dimensional image viewer for Python [Internet]. 2022. Available from: https://zenodo.org/record/6598542

75. Wang C, Dinesh RK, Qu Y, et al. CRISPRa screening with real world evidence identifies potassium channels as neuronal entry factors and druggable targets for SARS-CoV-2 [Internet]. bioRxiv. bioRxiv; 2021. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.07.01.450475

76. Chan K, Farias AG, Lee H, et al. Systematic genome-scale identification of host factors for SARS-CoV-2 infection across models yields a core single gene dependency; ACE2 [Internet]. bioRxiv. bioRxiv; 2021. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.06.28.450244

77. Blanco-Melo D, Nilsson-Payant BE, Liu W-C, et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. Cell. **2020**; 181(5):1036–1045.e9.

78. Li T, Kenney AD, Liu H, et al. SARS-CoV-2 Nsp14 activates NF-κB signaling and induces IL-8 upregulation. bioRxiv [Internet]. **2021**; . Available from: http://dx.doi.org/10.1101/2021.05.26.445787

79. Raj VS, Mou H, Smits SL, et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. Nature. **2013**; 495(7440):251–254.

80. Li W, Moore MJ, Vasilieva N, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature. **2003**; 426(6965):450–454.

81. Mimitou EP, Cheng A, Montalbano A, et al. Multiplexed detection of proteins,

transcriptomes, clonotypes and CRISPR perturbations in single cells. Nat Methods. **2019**; 16(5):409–412.

82. Simoneau CR, Chen PY, Xing GK, et al. NF-κB inhibitor alpha has a cross-variant role during SARS-CoV-2 infection in ACE2-overexpressing human airway organoids. bioRxiv [Preprint]. 2022 Aug 2:2022.08.02.502100.

Methods

Establishment and propagation of SARS-CoV-2 clinical isolate

SARS-CoV-2 (SARS-CoV-2/human/USA/CA-UCSF-0001H/2020) was isolated, propagated, and plaqued on Huh7.5.1 cells overexpressing ACE2 and TMPRSS2 [9]. Viral titer was determined using standard plaque assay with Avicel [68] on Huh7.5.1-ACE2-TMPRSS2 cells. Isolated virus was sequence-verified, lineage identified using PANGO [69], and deposited onto GISAID (accession ID: EPI_ISL_13689582). Additionally, SARS-CoV-2 was mycoplasma negative (Lonza MycoAlert Mycoplasma Detection Kit). All experiments in this study that utilized cultured SARS-CoV-2 were conducted in a biosafety-level 3 laboratory.

Cell Culture

The CRISPRi Calu-3 cell line was generated by lentiviral delivery of pMH0001 (UCOE-SFFV-dCas9-BFP-KRAB) [19] (Addgene #85969) into Calu-3 cells, followed by FACS sorting of BFP positive cells [19,35]. These cells were grown in DMEM/F12 supplemented with 10% FCS, penicillin, streptomycin, glutamine and non-essential amino acids. Huh7.5.1 cells overexpressing ACE2-TMPRSS2 and HEK293Ts cells were grown in DMEM supplemented with 10% FCS, penicillin, streptomycin, and glutamine. All cell types were maintained at 37°C and 5% CO_2.

Library design and lentivirus generation

Our Perturb-seq library was designed to target coronavirus host factors which were compiled from the literature, primarily from proteins physically interacting with coronavirus proteins, and from genes that came up as hits in CRISPR screens for host factors. All targets,

sgRNA sequences, and host factor annotations are listed in Table S1. Guide selection and library cloning followed the design introduced by Replogle et al [37,38]. We used a lentiviral backbone (pJR101, a variant of pJR85, Addgene #140095, with a GFP instead of BFP marker) which carries an additional Puromycin marker and allows the expression of two separate sgRNAs from different U6 promoters (human and mouse, respectively) with two distinct sgRNA constant regions (CR1 and CR3, respectively) to remove homologous regions in order to minimize recombination during lentiviral packaging. CR1/3 are further engineered with 'capture sequence 1' to be compatible with 10x's direct guide capture technology of the non-polyadenylated sgRNAs [37].

Guide oligos containing sets of two sgRNA sequences, separated by a spacer region, were ordered from Twist Bioscience, PCR-amplified, and cloned into pJR101 by ligation into the BstXI/BlpI restriction sites. The BsmBI-flanked spacer was then replaced by a fragment amplified from pJR98 (Addgene #140096), carrying the constant region of the first sgRNA and the promoter for the second one. The resulting library was sequenced to confirm proper guide sequences and abundance distribution.

After initial library cloning was completed, we obtained new screening data and designed an additional 24 sgRNAs, targeting 12 factors with 2 sgRNAs each. Those were cloned in array into the same pJR101 background as one-guide vectors (without the pJR89 drop-in). We then pooled the individually cloned sgRNA vectors with the initial library at equimolar amounts of all library elements at the DNA level. We used this combined library for lentiviral production as described [37]. While analyzing our single-cell datasets, we observed that the individually cloned library elements were overrepresented roughly 3-fold, which we attribute to higher lentiviral packaging efficiency due to their slightly smaller size.

Perturb-seq

Calu-3 CRISPRi cells were transduced with our Perturb-seq library at an MOI of ~0.1. Cells were puromycin-selected for 7 days, after which they had plateaued at ~93% GFP⁺ cells, followed by two more days of culture without selection markers. Cells were seeded into a 12-well plate at 400,000 cells/well and on the following day infected with SARS-CoV-2 at an MOI of 4. Infection was performed either for 1h, followed by a media change ('pulsed infection') or without removal of the inoculum ('non-pulse'). After 24 hours, cells were washed with PBS, dissociated with TrypLE Select Enzyme (10x, Thermofisher Scientific), washed, and resuspended in 1x PBS with 0.04% BSA. Wild-type, uninfected Calu-3 cells were spiked at ~1% into the dissociated Calu-3 CRISPRi cells to allow for analysis of ambient viral RNA. Manufacturer's instructions for the Chromium NextGEM Single Cell V(D)J Reagents Kit v1.1 (10x Genomics) were followed for preparation of gene expression libraries. Modifications to the 10x single-cell sequencing protocol were made for direct guide capturing and library preparation as previously described [37]. Gene expression and guide libraries were subsequently quantified on the Bioanalyer (Agilent) using the High Sensitivity DNA kit, pooled, and sequenced on the Illumina NovaSeq 6000 (read 1: 150 bp, read 2: 150 bp, index length: 8 bp).

Data analysis

Gene expression libraries were aligned using the 10x Genomics CellRanger v3.1.0 with default settings and aligned the hg38 reference genome concatenated with the SARS-CoV-2 genome. For viral alignments, STARsolo (version 2.7.8a) was used to capture viral junction sites. Cell barcode and UMI were identified for guide libraries using CellRanger. Guides identity was assigned to single cells following the Replogle et al. mixed model approach. Infection conditions

were combined for downstream analyses since there was not a statistically significant different distribution in guides between the conditions.

Scanpy was used for downstream cell filtering and analyses [70]. Cell filtering was done to include only cells that have one guide set per cell and at least 55 cells per guide. Additionally, low quality cells characterized as the bottom 2% of cells in total counts and cells with greater than 20% mitochondrial RNA were excluded.

We found the two populations with pulsed vs non-pulsed infection to exhibit very similar characteristics and combined them for all downstream bioinformatic analyses.

For assessing the effect of the host transcriptome in the setting of vastly different library distributions, we performed experiments with and without viral sequences, and subsequently downsampled to the level of infected cells and re-analyzed the data.

For clear identification of infected cells, ambient viral RNA was evaluated in wild-type Calu-3 cells packaged into droplets. These WT control cells were identified by selecting cells that lack Cas9, lentiviral, and guide transcripts. We additionally selected cells that have at least 10,000 UMIs, which yielded 1082 cells for this analysis. In those cells, we determined the mean and standard deviation of the read counts of all individual viral genes. Other cells were considered infected if they had at least 6 viral transcripts at 2 standard deviations above the mean of WT cells, as well as more than twice the total viral transcript reads per cell. Conversely, cells were considered uninfected if no viral gene exceeded the 2 standard deviation threshold. A small proportion of cells could not be clearly determined as infected or uninfected, therefore we classified these cells as the borderline population (designated cluster T in UMAP space).

Guide knockdown percentages were determined by calculating the normalized count of target gene / nontargeting control. This analysis was limited to bystander cells to remove the effects

of viral antagonism, and was subset to genes with at least 0.5 UMI per cell after normalization to remove low abundance or undetectable genes. Cell cycle phases were determined following scanpy's tutorial. Similarly, single cells were also scored for interferon stimulation (ISG score) using scanpy's sc.tl.score_gene function. Differential expression was performed by exporting scanpy's count matrix to R, and subsequently performing MAST following Seurat's tutorials.

Orthogonal Validation

For targeted follow-up, published protocols for guide design and cloning into the lentiCRISPR v2 plasmid were followed [71]. The following sgRNA sequences were used:

ACE2: CACCGCAGGATCCTTATGTGCACAA

NFKBIA: CACCGAGGCTAAGTGTAGACACGTG (Huh7.5.1),

CACCGCTGGACGACCGCCACGACAG (HEK293T)

EIF4H: CACCGCCCCCCTACACAGCATACGT

EIF4E2:CACCGTCATAGCTCTGTGAGCTCGT

Lentivirus was produced in HEK293Ts by co-transfecting pMD2.G, DR8.91, and the lentiCRISPR v2 plasmid with the guide of interest using TransIT-Lenti (Mirus Bio). Lentivirus-containing supernatant was collected 48 hours after transfection, filtered, and frozen.

For orthogonal validation of host factors that alter viral infection, Huh7.5.1-ACE2-TMPRSS2 cells were transduced with lentivirus in the presence of polybrene. We next selected transduced cells for 72 hours with puromycin. After 1 week, knockout cell lines were infected with SARS-CoV-2 at an MOI of 3 for 20 hours in biological duplicates. Cells were subsequently fixed in 4% paraformaldehyde for 30 minutes, permeabilized with 0.2% Triton X, blocked with 5% BSA stained with primary anti-NP (Sino Biological 40143-R001), and secondary goat anti-rabbit IgG

conjugated to Alexa Fluor 488 (Thermo Fisher Scientific A-11034). Slides were mounted with DAPI Fluoromount-G (SouthernBiotech 0100-20) and imaged on a Nikon Ti inverted fluorescence microscope (4X). Quantification of images was performed using CellProfiler 4 [72].

To investigate the effect of NFKBIA knockout on NF-κB induction, we generated NFKBIA knockout lines and controls in a background of HEK293T cells expressing N-terminally mNG11-tagged RELA [45]. RELA-tagged cell lines were transduced with lentivirus carrying Cas9 and NFKBIA-targeting sgRNA and puromycin-selected for 1 week. Cells were stimulated with recombinant TNF-α (50 ng/ml; Abcam ab9642) and imaged using confocal microscopy 25 and 45 minutes after stimulation. The imaging volume per field of view was 21 μm depth with 0.25 μm z-sectioning. During imaging, cells were maintained in a stage-top incubator (Okolab, H201-K-Frame) at 37°C and 5% CO2. The imaging was performed using a DMI-8 inverted microscope (Leica) with a Dragonfly spinning-disk confocal (Andor) with a 63x 1.47 NA oil objective (Leica). Images were acquired using a Prime BSI sCMOS camera (Photometrics, pixel size = 6.5 μm x μm). Microscope control was achieved with Micromanager version 2.0.0 [73]. Image visualization was via napari v0.4.16 [74].

Author Contributions

S.S., J.L.D. J.S.W. and M.Y.H. conceptualized the study.

M.Y.H., S.S., A.S.P. and J.M.R. designed the Perturb-seq library with input from M.B.F.

A.S.P. and M.Y.H. cloned the library.

J.K.N. engineered Calu-3 CRISPRi cells.

S.S. performed live-virus, single-cell and follow-up experiments with help from M.Y.H., B.S.Z., J.L. and M.T.L.

O.S.R. oversaw BSL-3 work.

M.D.L. provided the mNG-RELA cell line and microscopy infrastructure.

S.S., M.Y.H., J.R.B. and A.H.M. performed microscopy follow-up work on RELA translocation.

S.S. and M.Y.H. analyzed the data with input from X.Q., C.J.Y. and J.W.

S.S. and M.Y.H. wrote the manuscript, J.L.D., J.S.W. and J.W. edited the manuscript. All authors commented on the manuscript.

Data Availability

Raw and preprocessed data are available on GEO (GSE208240).

Analyses will be made public on Github with publication.
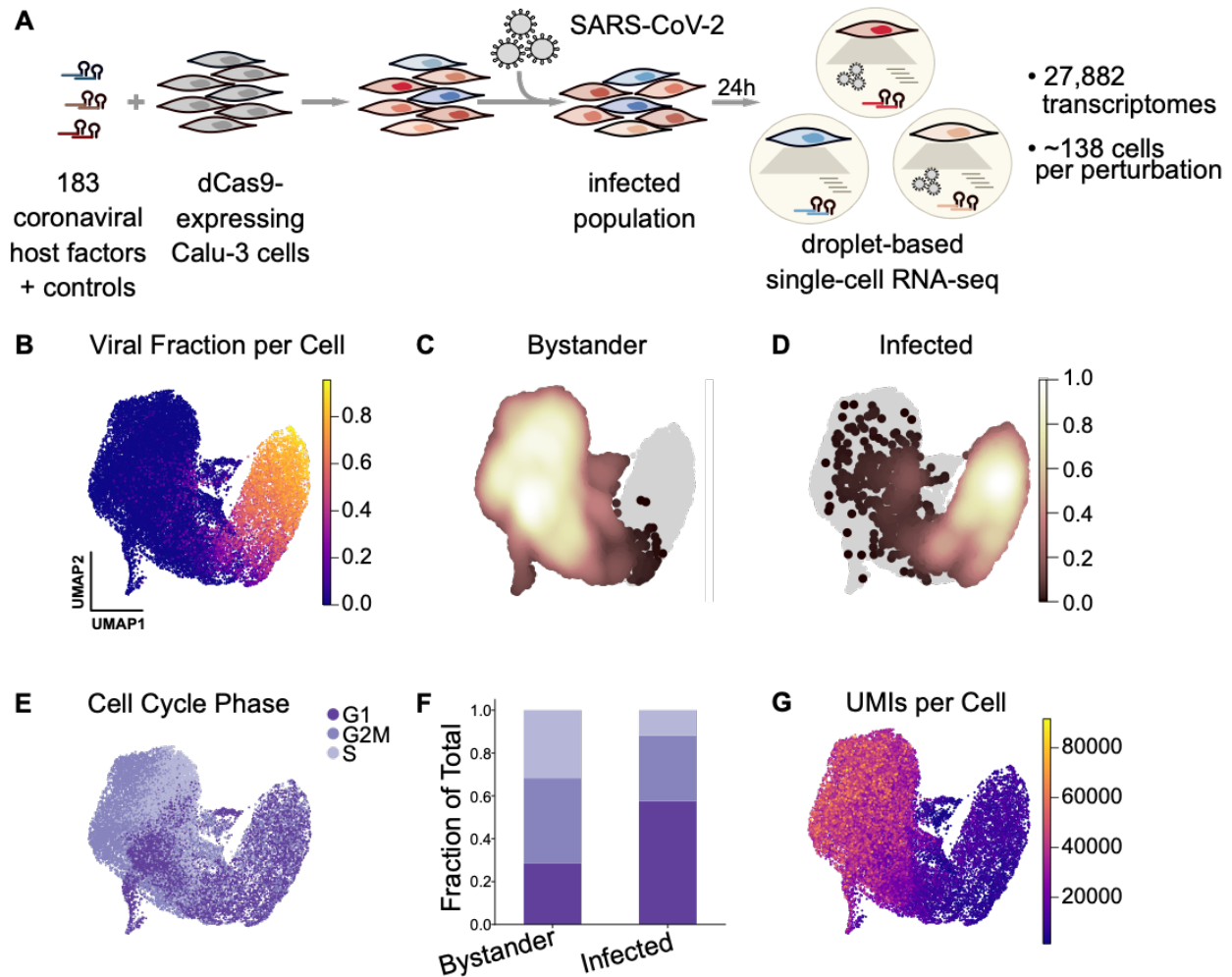
Figure 6.1: Perturb-seq for single-cell transcriptional analysis and functional validation of SARS-CoV-2 host factors. A. Experimental design for the Perturb-seq experiment in Calu-3 cells engineered to express CRISPRi machinery. We perturbed 183 different host factors (individually or in combination) using a lentivirally-delivered library, infected the cells with SARS-CoV-2 for 24 hours, and performed droplet-based single-cell RNA sequencing, reading out host and viral transcripts as well as the sgRNA, indicating the perturbed host factor. **B.** Single-cell transcriptomes were projected into UMAP space and colored by viral RNA fraction per cell. **C, D.** Density of cells identified as either uninfected/bystander (C) or infected (D) by our classifier, overlaid onto all cells in gray. **E.** Cells color-coded by their cell cycle phase. **F.** Fraction of bystander and infected cells assigned to each cell cycle phase. **G.** Cells color-coded by the number of detected UMIs per cell.
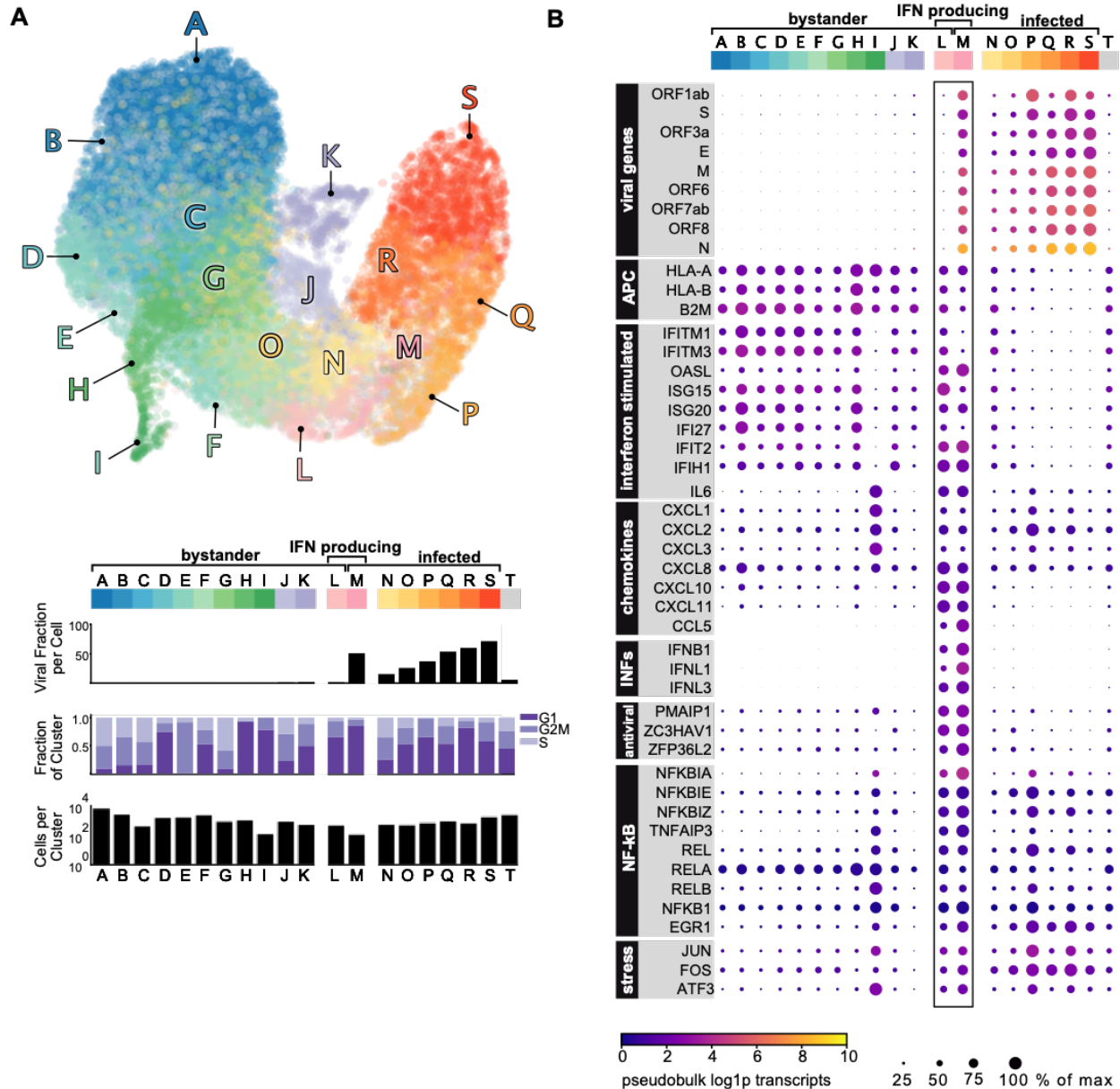
Figure 6.2: Transcriptional heterogeneity in SARS-CoV-2 infection. A. Single-cell transcriptomes were projected in UMAP space and colored by Leiden cluster. Leiden clusters were subsequently characterized by the mean viral fraction, the number of cells, and the cell cycle composition per cluster. Cluster T are all cells that could not be assigned an unambiguous infection state. **B**. Differential expression of Leiden clusters revealed transcriptionally distinct subclusters of bystander cells, infected cells, and a small subset of interferon-producing cells. The color of each dot is pseudobulk gene expression of each gene per cluster, and the size of each dot is the expression normalized to the cluster with maximum expression of that gene.
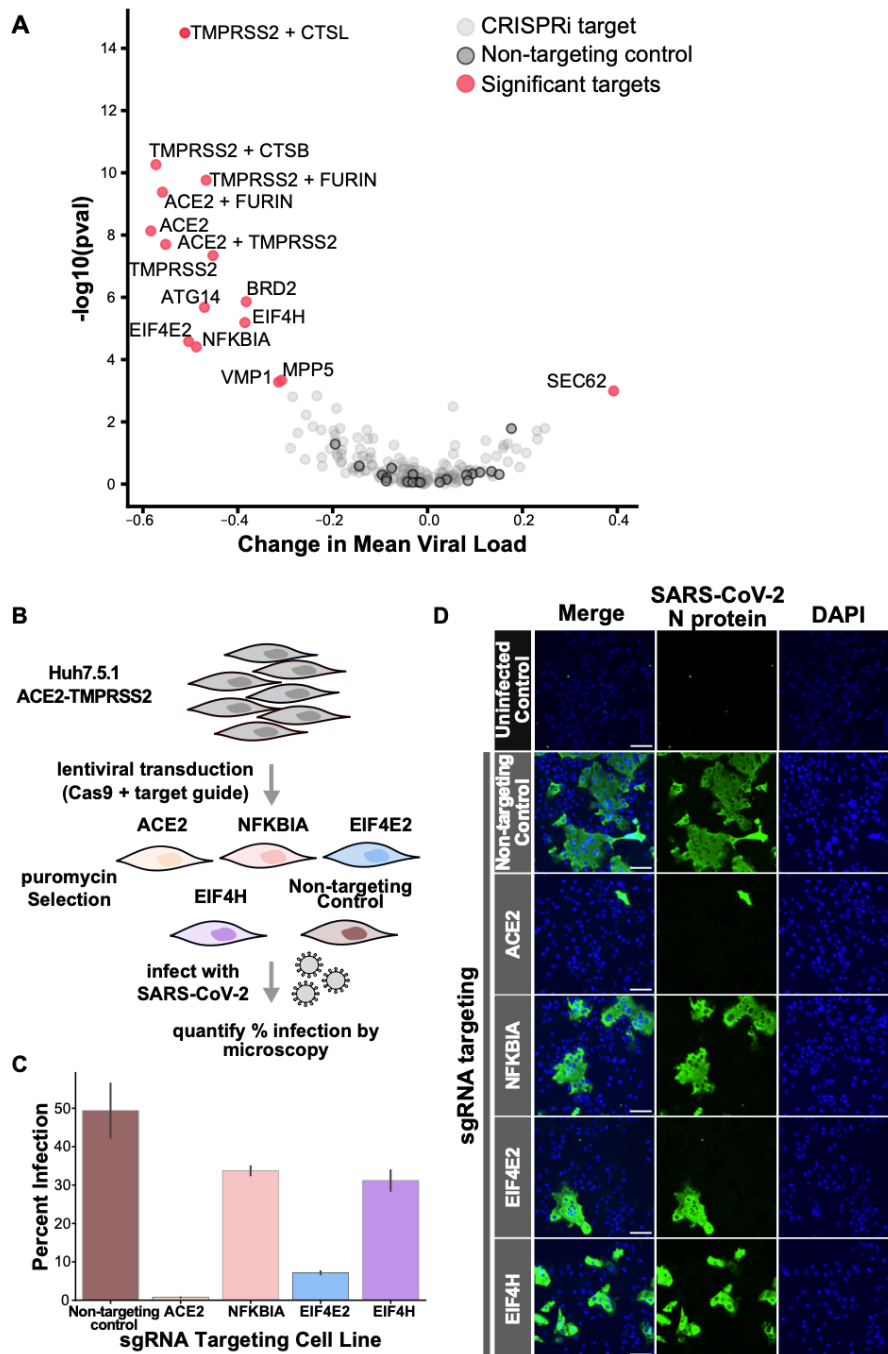
Figure 6.3: Host perturbations alter SARS-CoV-2 infection dynamics. A. The effect of how each CRISPRi perturbation altered viral load was displayed as the change in mean viral load by KS p-value of viral load distribution change. Color code indicates CRISPRi targets, non-targeting controls and targets in which knockdown significantly altered viral loads. **B.** To orthogonally validate CRISPRi targets, we transduced Huh7.5.1 cells overexpressing ACE2 and TMPRSS2 with lentivirus targeting control and test genes. Cells were subsequently infected with SARS-CoV-2, and percent infection was calculated (**C**) by immunofluorescence and microscopy (**D**). Scale = 100 μm.
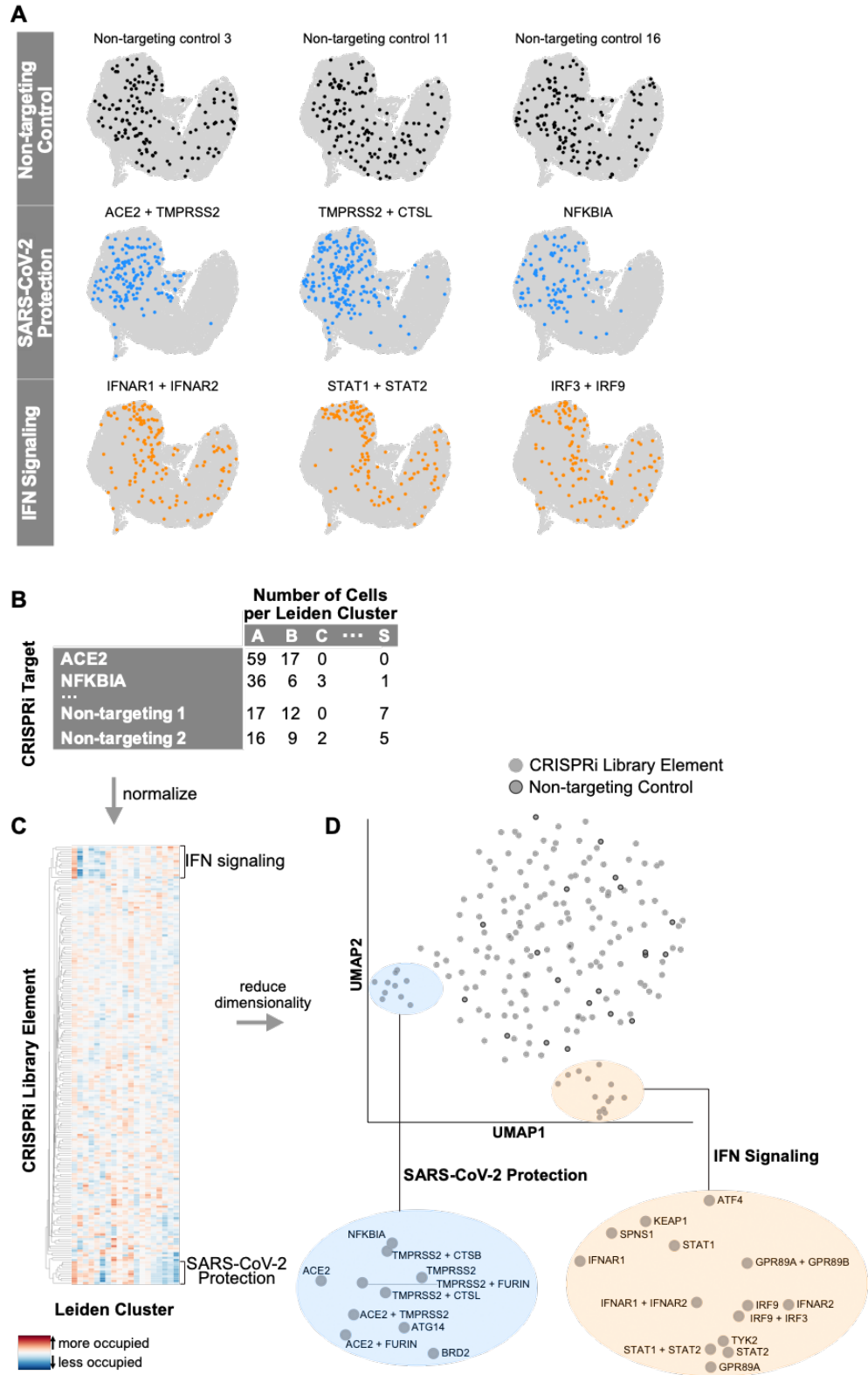
Figure 6.4: Host perturbations alter localization of cells in UMAP space and Leiden cluster membership. A. Library elements for non-targeting controls, factors that alter SARS-CoV-2

infection, and interferon signaling are highlighted in UMAP space. **B.** Library element representation by cluster was calculated, normalized and visualized on a heatmap **(C). D.** Subsequent dimensionality reduction of this odds-ratio was projected into UMAP space and revealed subclusters by biological function.
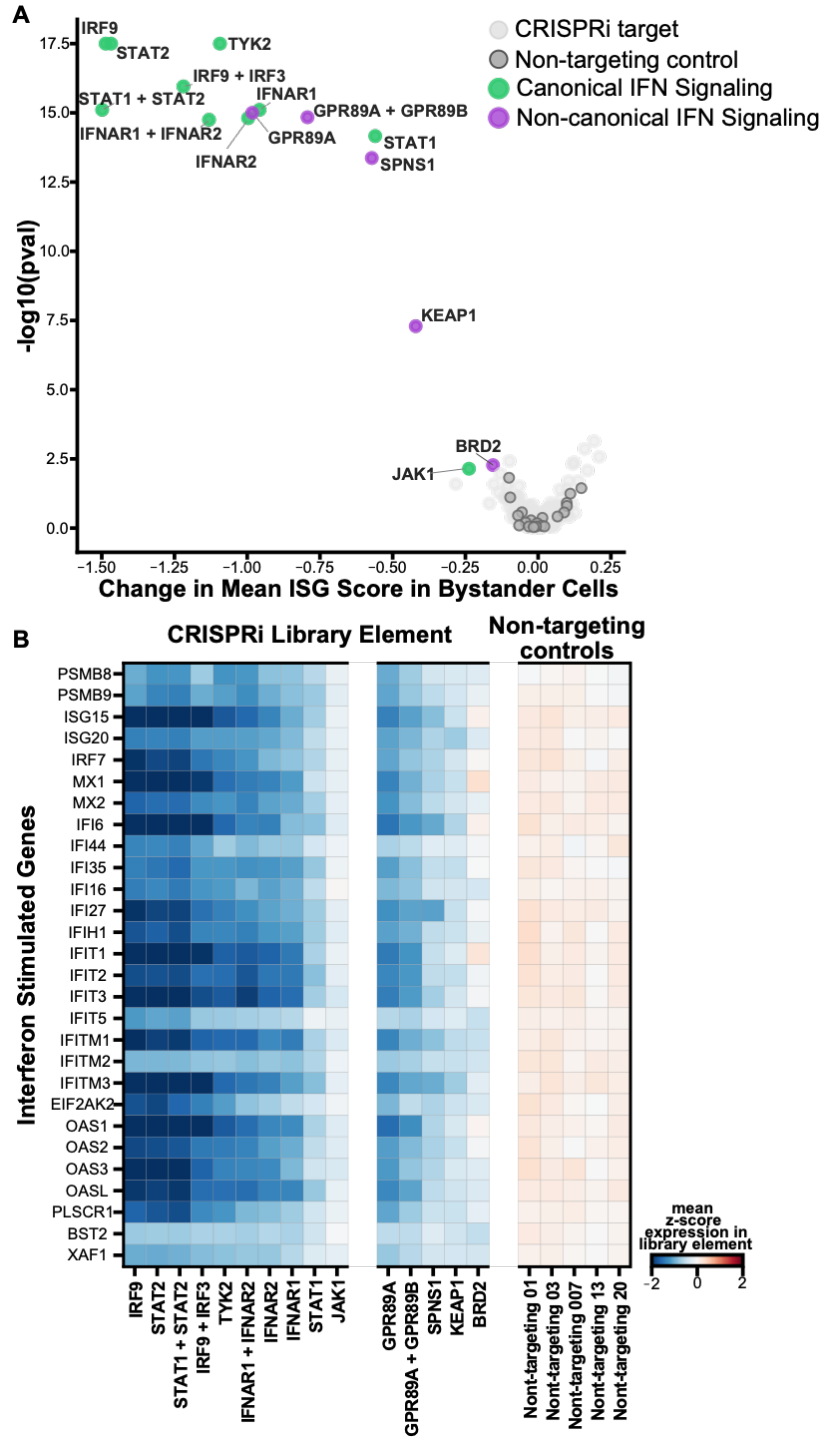
Figure 6.5: Host perturbations alter interferon signaling in bystander-activated cells.
A. We scored bystander cells based on their ability to respond to interferon (ISG score) and tested which perturbations significantly altered the ISG score distribution by perturbation. This is represented by the mean change in ISG score when compared to non-targeting controls by the KS p-value per perturbation. **B.** Expression heatmap of select targeting and non-targeting library elements showing the mean z-score for a subset of interferon-stimulated genes.
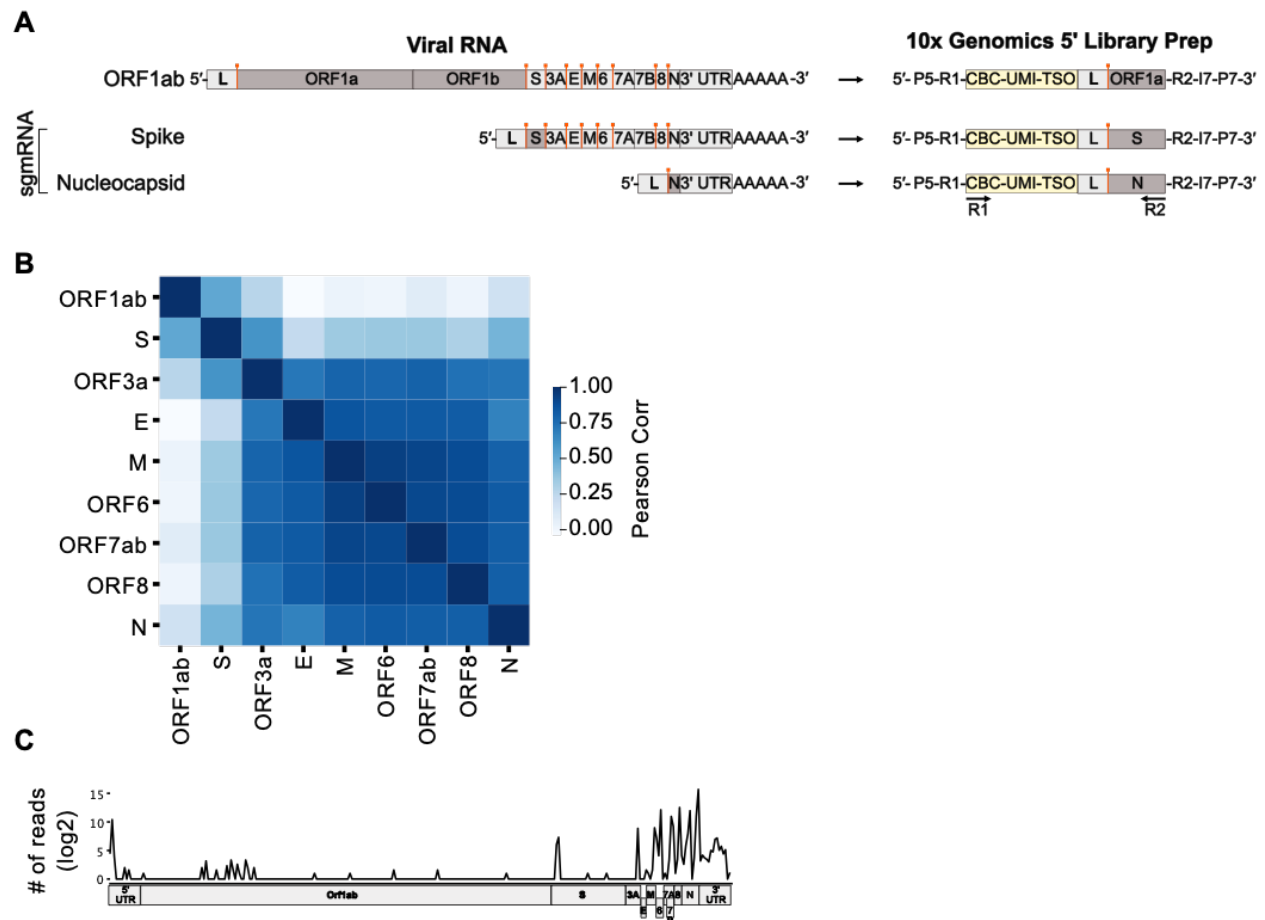
Figure 6.6: Characterization of viral transcripts. A. Schematic representing the structure of SARS-CoV-2 viral RNA, and the expected sequence that captures the lead-body junction (represented with orange bar) using the 10x Genomics 5' workflow. **B.** Pearson correlation matrix of all viral RNA transcripts. **C.** Viral reads from infected cells were extracted and junction sites were mapped to the viral genome.
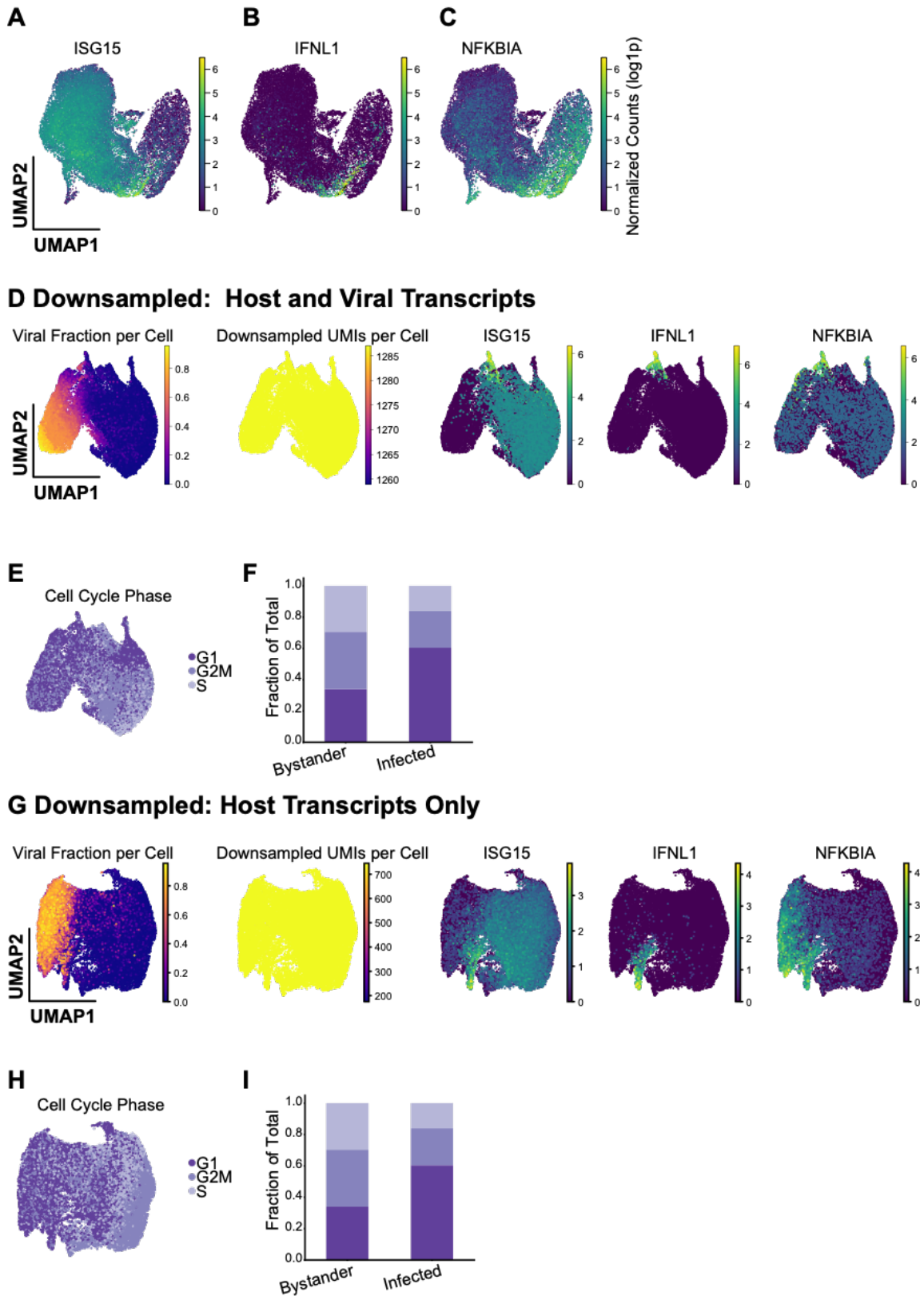
Figure 6.7: Host gene expression and downsampling show similar transcriptional phenotypes. Host transcriptional analysis revealed heterogeneity in infected and bystander

populations, including differential gene expression in UMAP space of: **A.** ISG15; **B.** IFNL1 and; **C.** NFKBIA. To confirm that transcriptional differences are not exclusively due to host shutoff, we downsampled host and viral transcripts and confirmed similar **D.** transcriptional patterns and; **E-F.** cell cycle phase patterns. Additionally, we removed viral transcripts and performed the same analysis to confirm observed phenotypes are not an artifact of including viral reads in analyses (**G-I**).
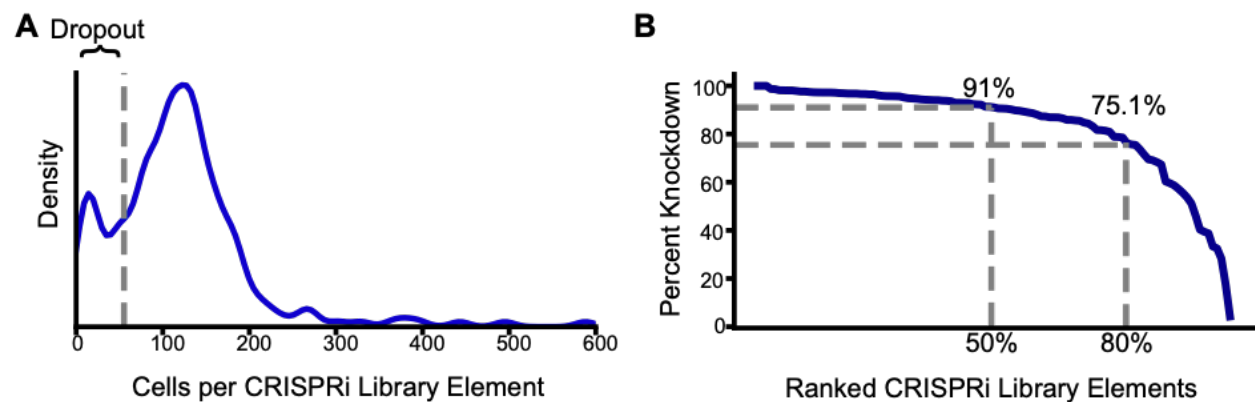
Figure 6.8: CRISPRi library element representation and knockdown efficiency. A. We evaluated the distribution of captured cells for our 239 library elements and plotted the kernel density estimate. **B**. Knockdown percentage for each element was calculated relative to non-targeting controls and ranked by percent knockdown.
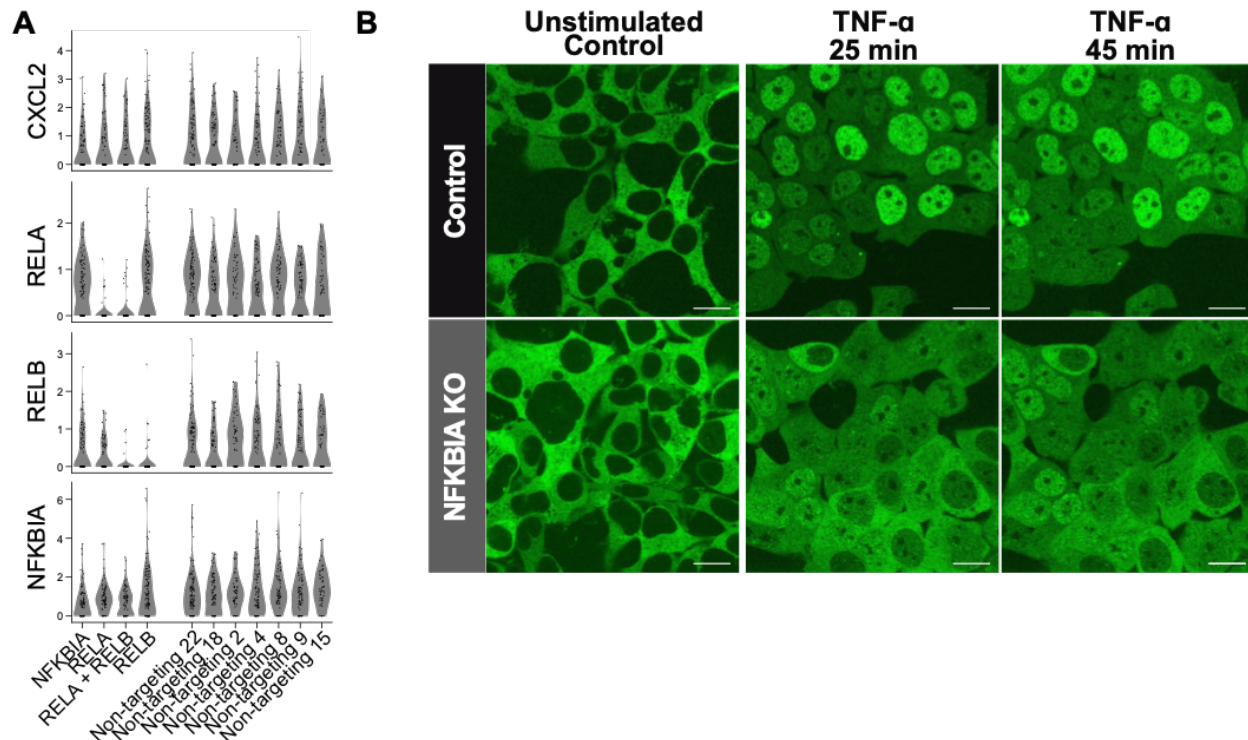
Figure 6.9: Effect of NFKBIA perturbation on transcription and p65/RELA localization. A. To evaluate if NFKBIA knockdown transcriptionally alters NF-κB signaling in bystander cells, we looked at gene expression (log1p) of NF-κB target genes (CXCL2) as well as the pathway members themselves, which are on a negative feedback loop (RELA/B, NFKBIA) in cells with guides targeting NFKBIA, as well as RELA, RELB, both RELA and RELB, and non-targeting controls. **B.** To further investigate the effect of NFKBIA perturbation on p65/RELA localization, we utilized split mNeonGreen (mNG)-tagged RELA cells, generated polyclonal CRISPR NFKBIA knockout cells, and monitored p65/RELA localization at baseline. Additionally, we stimulated genetically unperturbed and NFKBIA KO cells with TNF-α and performed live-cell imaging at 25 and 45 minutes after stimulation (scale = 20 μm).

Figure 6.10: Localization of CRISPRi library elements in Leiden clusters (expanded Figure 6.4). Expanded version of heatmap and UMAP with all CRISPRi elements labeled in A. and B.

Chapter 7: Identification of Host Factors and Characterization of the Antiviral Response to

RSV Infection

The work in this chapter includes contributions from:

Dr. Andreas Puschnik, Dr. Hanna Retallack, Dr. Matthew Laurie, Dr. Jamin Liu, Dr. Duo Peng,

Dr. Kristeene Knopp, Dr. Matthew Zinter, Dr. Chun Jimmie Ye and Dr. Joseph DeRisi.

Abstract

Respiratory syncytial virus (RSV) is a prevalent pathogen globally, and remains the leading cause of bronchiolitis and pneumonia in the United States for children during their first year of life [1]. Despite its prevalence worldwide, there remains no efficacious therapeutic for infected patients. To facilitate investigation of the host response to this common respiratory pathogen, and to aid in the discovery of host determinants of RSV infection, we leveraged two unbiased technologies: single-cell RNA sequencing and whole genome CRISPR/Cas9 knockout screening. Our unique data complement prior studies that investigate the proinflammatory response to RSV infection, and provide a rich resource to aid in future hypothesis testing.

Introduction

Respiratory Syncytial Virus (RSV) is a ubiquitous respiratory virus that infects most children by two years of age, and reinfections are common throughout life [2]. RSV can lead to both upper and lower respiratory tract symptoms for all age groups, but remains a salient cause of infant mortality. Despite numerous clinical trials, there remains no RSV-specific therapeutic for infected patients. Due to the global health burden of RSV infection and lack of efficacious therapeutics, it remains essential to continue investigating the fundamental molecular virology of infection and the host response to RSV.

RSV is a negative-sense, single-stranded RNA virus, and a member of the *Paramyxoviridae* family [3]. It primarily infects respiratory epithelial cells, and a multitude of bulk transcriptional and proteomic studies have investigated the host response to RSV infection [4–7]. Upon infection, these respiratory epithelial cells sense viral RNA (RIG-I, MDA-5, TLRs) and initiate proinflammatory signaling cascades (interferon and NF-κB signaling) [8]. The strength

118

and timing of the innate immune response in the respiratory epithelium is crucial for effective viral clearance. Notably, to evade the host antiviral response, RSV encodes two nonstructural proteins (NS1 and NS2) that inhibit both interferon induction and signaling [6,9,10]. A detailed understanding of critical host pathways activated and repressed during infection, and pathways altered by RSV antagonism remains critical for novel therapeutic discovery.

While the host response to RSV infection has been investigated *in vitro, ex vivo and in vivo* using transcriptomics and proteomics [4–7], it has remained challenging to understand the complexity of the host response using these bulk methods. Over the last decade, single-cell resolution has led to the appreciation of the heterogeneous response to many viral infections [11–15], and has revealed that both infected and bystander activated cells contribute to the proinflammatory response. In the setting of RSV infection, the relative contribution of infected and bystander cells to the proinflammatory response remains undetermined.

In addition to investigating the transcriptional response to viral infections, perturbation based screens are commonly used to identify host dependency factors of viral infection. While both RNAi [16] and haploid [17] screens have effectively unveiled host dependencies of RSV infection, CRISPR/Cas9 screening may further complement these studies by genetically knocking out all host genes in respiratory epithelial cell lines.

Here, we utilized two unbiased approaches to investigate pathways altered by and necessary for RSV infection in a respiratory epithelial cell line: droplet-based single-cell RNA sequencing and whole genome CRISPR/Cas9 screening. Our results identified host factors and pathways that are differentially expressed in infected and bystander activated cells, and revealed host dependencies factors of RSV.

Results

Single-cell transcriptomics reveals varying cellular states of RSV infection

To characterize the transcriptional response of RSV infected and bystander activated cells over the viral life cycle, we performed temporal single-cell RNA sequencing. We infected a human adenocarcinoma lung epithelial cell line (A549 cells) with RSV (subtype A2), and performed droplet based single-cell RNA sequencing (10x Genomics) at four time points: 0 hrs, 4 hrs, 8 hrs and 12 hrs post infection (Figure 7.1A). Additionally, to investigate differences between cells that underwent true infection compared to cells that came into extracellular contact with viral proteins and RNA, we treated A549 cells with heat inactivated RSV for each of the specified time points.

First, we created a classifier for identifying infected and bystander activated cells as described in Sunshine et al. [18]. In brief, to assess the presence of ambient viral RNA from the cellular supernatant, we included a spike-in of murine 3T3 cells prior to single-cell RNA sequencing to allow for computational identification of infected versus bystander cells. With this classifier, we were able to identify RSV infected populations and calculated the percent infection by condition and time point (Figure 7.1B). With an MOI 0.3, we observed 28.6% infection by 12 hours post inoculation. Both uninfected and heat-inactivated controls remained uninfected for the duration of our experiments, therefore we are confident with both the experimental protocol and classifier utilized in these experiments.

We next sought to understand the transcriptional characteristics of the infected population. Across all infected cells, we observed a mean viral fraction per cell of 8.1% and maximum of 40.3%. We subsequently investigated how the viral fraction per cell changed over time, and by condition (Figure 7.1C). Over our time course, we saw a steady increase in the number of cells

with a higher viral fraction, suggesting a steady increase in viral load per cell over this 12 hour time course.

Additionally, we observed that cells with a higher fraction of viral RNA had more total RNA counts (UMIs) than cells with less or no viral RNA (Figure 7.1D). Upon removal of viral reads, this distribution shifts down near the uninfected/lowly infected population (Figure 7.1E). This indicates that unlike other viral infections (i.e. SARS-CoV-2), RSV does not lead to host shutdown during infection.

To further understand transcriptional patterns within our infected population, we next tested the correlation of every host gene with the most abundant viral transcript, NS1. Pearson's correlation revealed 8 human genes that correlated (r >0.3) with viral NS1 (Table 7.1): HERPUD1, HSPA5, MANF, SDF2L1, HSPA1A, DDIT3, DNAJB9, HSPE. Notably, many of these host factors are involved in regulation of endoplasmic reticulum stress and the unfolded protein response.

To initially characterize the transcriptional heterogeneity of RSV infection, we subsetted our data to the time point that had the highest abundance of infected cells (12 hours). At this time point, our analysis revealed close overlap in UMAP space of our cells treated with the vehicle control and heat inactivated RSV (Figure 7.2A). On the contrary, cells that were treated with live RSV are distributed throughout UMAP space with the highest density in two major clusters. Further investigation of these clusters revealed a cluster of infected cells, with varying fraction of viral RNA and viral transcripts (Figure 7.2B-D). Both correlation analysis and differential expression revealed an upregulation of ER stress genes HERPUD1 and DDIT3 within infected cells (Figure 7.2E-F). The second major cluster identified cells lacked viral RNA, have an

increased abundance of ISG15 (Figure 7.2G) and other interferon stimulated genes (ISGs) such as MX1, IFIT1, and NT5C3A.

ISG15 is a well described ISG with known antiviral functionality [19]. Along with being upregulated in our bystander population, it is additionally upregulated in bystander activated cells in the setting of SARS-CoV-2, HCMV and influenza infections [11,14,18]. To further identify other genes that are co-expressed with ISG15 in our bystander population, we subsequently performed Pearson's correlation of ISG15 with all other host genes (Table 7.2). This analysis confirmed co-expression of validated ISGs within our bystander activated population, and complements prior studies [14] in describing ISGs that are detectable by single-cell RNA sequencing.

To further investigate gene expression differences between each condition and time point, we next performed differential expression for A549 cells treated with RSV and heat inactivated virus, compared to uninfected controls. There is not a detectable difference in signal at early time points between heat inactivated and vehicle treated controls. We observed few genes differentially expressed at early time points, and began to see an upregulation of ISGs by 8 hours post infection. Differential expression (MAST [20]) of infected versus bystander activated cells revealed a number of pathways 12 hours post infection (Figure 7.2H). We observe a downregulation of multiple ISGs (ISG15, IFIT1, IFI6, STAT1, MX1) and CYP24A1 (cytochrome P450 enzyme family), in infected cells compared to the bystander activated cells. Furthermore, we see an upregulation in ER related (HERPUD1, SELENOS, SDF2L1, MANF) and plasminogen related (PLAU, PLAUR) genes. We additionally see an upregulation of apoptotic factor, TNFRSF12A, and lysosomal protease, CTSL in RSV infected cells.

Genome-wide CRISPR Screen Identifies Host Dependency Factors of RSV Infection

To identify host determinants of RSV infection, we performed whole-genome CRISPR/Cas9 knockout screening. First, A549 cells were transduced with the GeCKOv2 library, and infected with RSV-EGFP for 24 hours. To enrich for host dependency factors, we performed two different iterations of the screen. For the first iteration, EGFP negative cells were sorted and cultured for 1 week, thus enriching for cells that were not infected at the time of the sort and remained viable after one week in culture. For the second iteration of the screen, EGFP negative and EGFP-low cells were harvested immediately after the sort (24 hours post infection). For both screens, sgRNAs that provided protection from RSV infection were PCR amplified and sequenced. We subsequently performed enrichment analysis using the MAGeCK algorithm [21] to rank genes and calculate the MAGeCK enrichment score for each screen (Figure 7.3A-C) .

Our viability screen (Screen #1) identified genes involved in heparan sulfate biosynthesis, the COG Complex and retrograde transport at the trans-golgi network (Fig 7.3B). Notably, the top hit for both screens was TMEM165 (Transmembrane Protein 165). This protein is localized to the Golgi apparatus (Human Protein Atlas [22]), and is functionally thought to play a role in protein glycosylation.

For the second iteration of our screen, along with heparan sulfate and COG complex related genes, we identified host factors related to the Golgi (ARFRP1, SYS1), V-ATPase (ATP6V1B2), vesicle trafficking (TRAPPC1) and membrane trafficking (RAB4A) (Fig 7.3B). We subsequently compared the two screens and saw substantial concordance in our top ranked genes (Fig 7.4A). We suspect the observed differences between the two screens is due to the viability screen (#1) enriching for sgRNAs that prevent infection and prolong cell life/proliferation, while our second

sort screen only enriches for sgRNAs that prevent infection within the first 24 hours. Overall, these screens highlight the importance of performing screens under different technical conditions.

Next, to assess if our top enriched genes are specific to RSV, we concatenated the top twenty genes of each screen, removed redundancy, and compared our findings to hits from prior RSV haploid and RNAi screens [16,17]. Additionally, we compared our findings to the hits of other viral genome-wide CRISPR knockout screens [23,24]. We observed 70% of our top genes had been described as proviral factors in these other viral studies (Fig 7.4B). This suggests that while many of these host factors altered RSV infection dynamics within these two screens, they may have pan-viral activity.

To further identify pathways and protein interaction networks for our list of enriched genes, we subsequently performed gene ontology, pathway and process enrichment analysis using Metascape [25]. Significantly enriched pathways include heparan sulfate and glycosaminoglycan biosynthesis, cytosolic transport, retrograde transport at the transit-golgi-network and endoplasmic reticulum to Golgi vesicle-mediated transport (Fig 7.5A). Furthermore, Metascape protein-protein interaction enrichment analysis identified two networks using the Molecular Complex Detection (MCODE) algorithm encompassing retrograde transport and heparan sulfate proteoglycan biosynthesis (Fig 7.5A).

Discussion/Future Directions

In this study, we used two unbiased approaches to investigate the single-cell response to RSV infection and uncover host determinants of infection. Our single-cell transcriptional analyses revealed upregulation of ISGs in bystander actuated cells, and this signature is largely absent within the infected population. We additionally observed an upregulation of genes associated with

ER stress in RSV infected cells. Notably, the signatures observed in the infected population do not mirror that of SARS-CoV-2 infected cells described in Chapter 6. We do not detect robust activation of the NF-κB inhibitors (NFKBIA/E/Z) within the RSV infected population, nor do we see host shutoff. Future directions include a deeper investigation into the dynamics and spatial landscape of interferon stimulation during RSV infection. More specifically, generation of ISRE-mCherry reporter lines, infection with RSV-EGFP and temporal live microscopy would allow for a spatial understanding of infection and the host response. Furthermore, a secondary question that comes from this work: how does perturbation of bystander activation alter infection dynamics? Future studies that investigate the functional consequences of impaired bystander activation may include: 1) mixing wild-type cells and IFNAR1/2 knockout cells at varying proportions; 2) infecting with EGFP; 3) performing temporal live microscopy. I hypothesize that perturbation of bystander activation will accelerate RSV infection.

We further sought to complement this rich transcriptional dataset and identify host determinants of RSV infection with genome-wide CRISPR screening. These analyses revealed multiple factors involved in heparan sulfate biosynthesis and golgi transport. Future directions include curating a list of genes from both the transcriptional and screening data, and similar to Chapter 6, performing Perturb-seq to functionally characterize and validate each gene's role in infected and bystander cells.

In summary, these datasets provide a resource for the broader RSV community, and lay the groundwork for further investigation of the host response and host determinants of RSV infection.
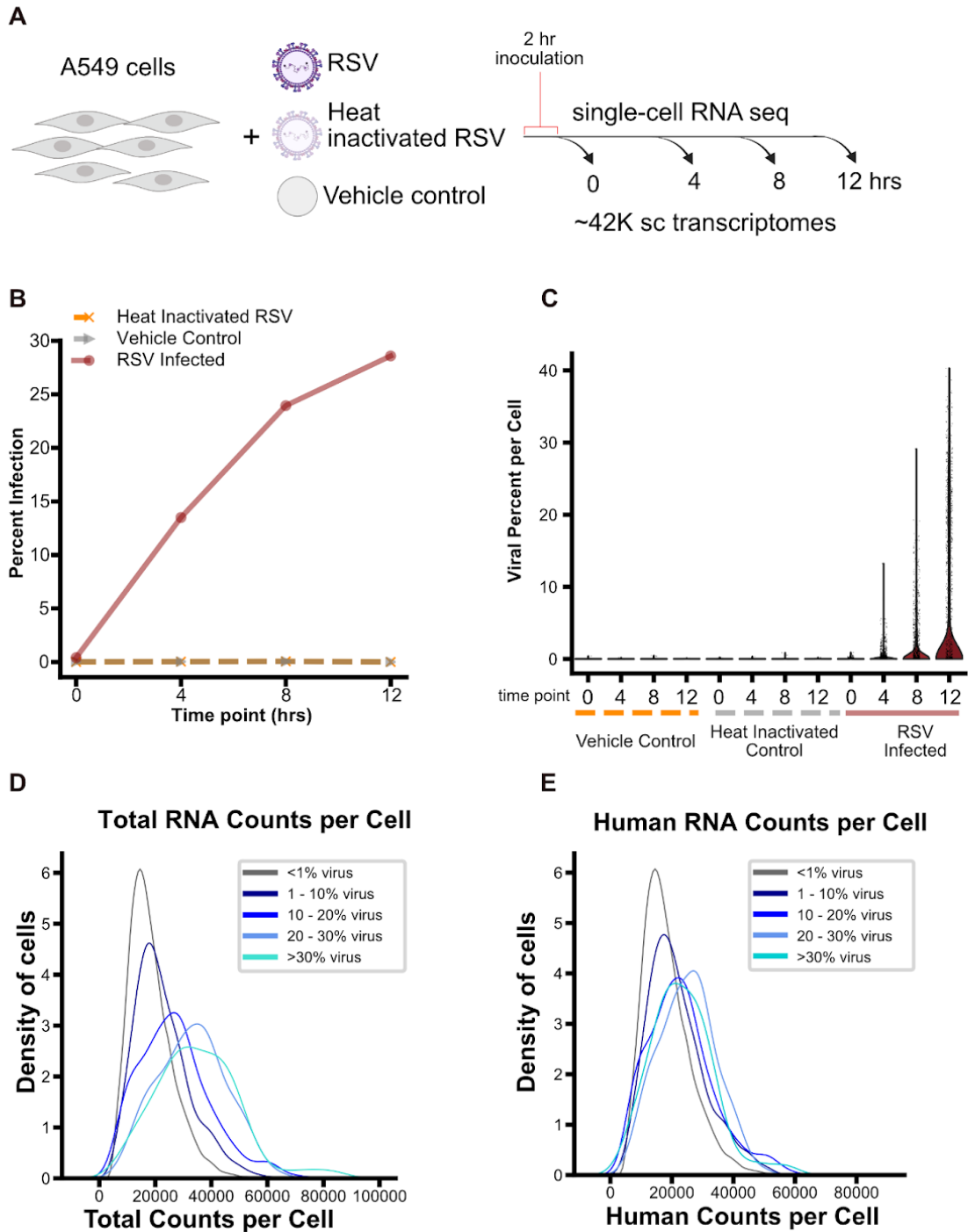
Figure 7.1: Experimental design and infection characterization. A. Schematic of experimental design. **B.** Percent infected broken down by time point and condition; **C.** Viral

percentage per cell for each condition and time point. All cells were binned by viral percentage per cell and each plot represents the density for (**D**) total counts and (**E**) human counts per cell.
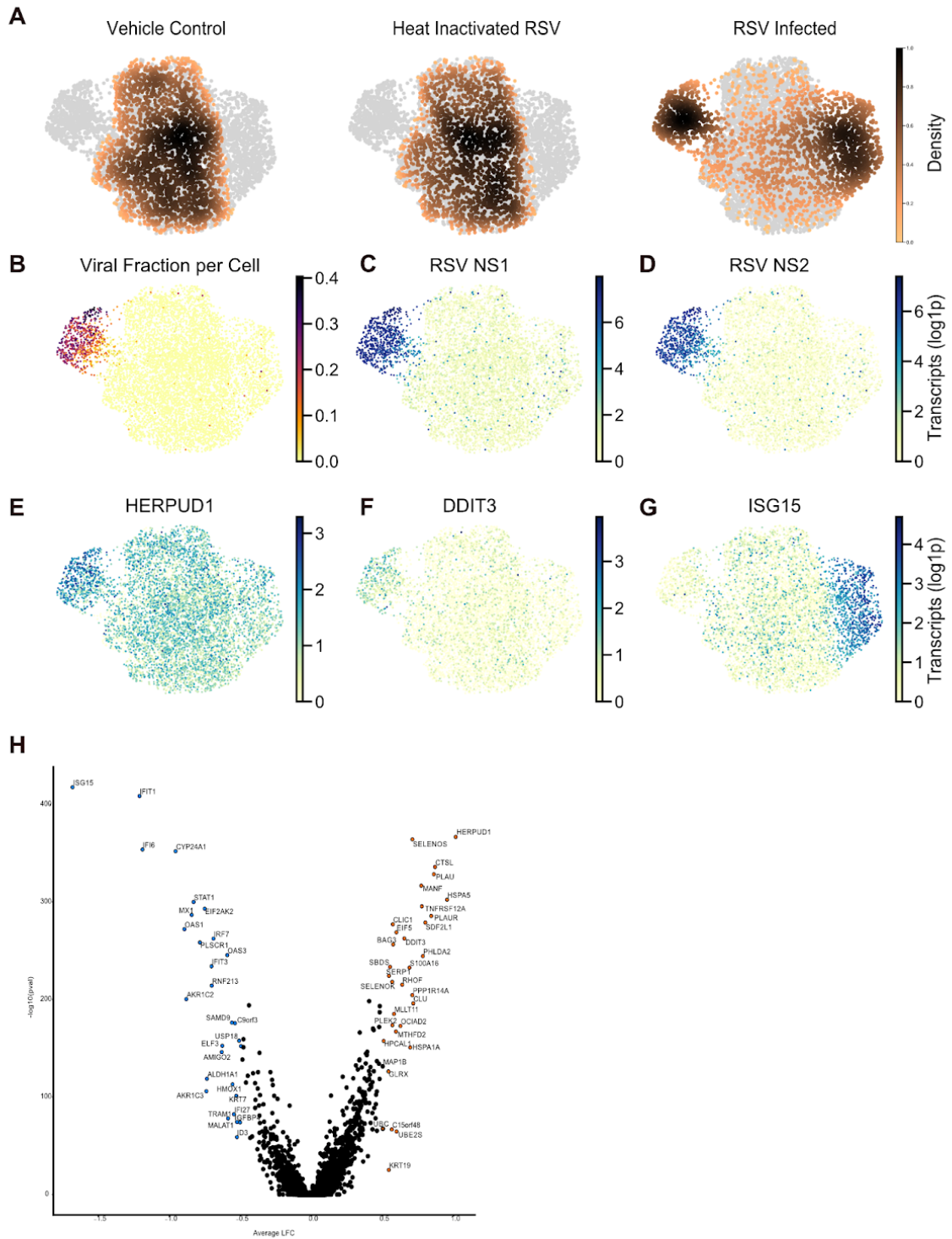
Figure 7.2: Single-cell transcriptome characterization at 12 hour time point. A. Density plot of different treatments. **B.** Viral fraction per cell. Viral transcript counts (log1p) of **C.** RSV NS1, **D.** RSV NS2, **E.** HERPUD1, **F.** DDIT3, **G.** ISG15. **H.** Volcano plot that highlights differentially expressed genes when comparing infected and bystander activated cells at 12 hours post infection.
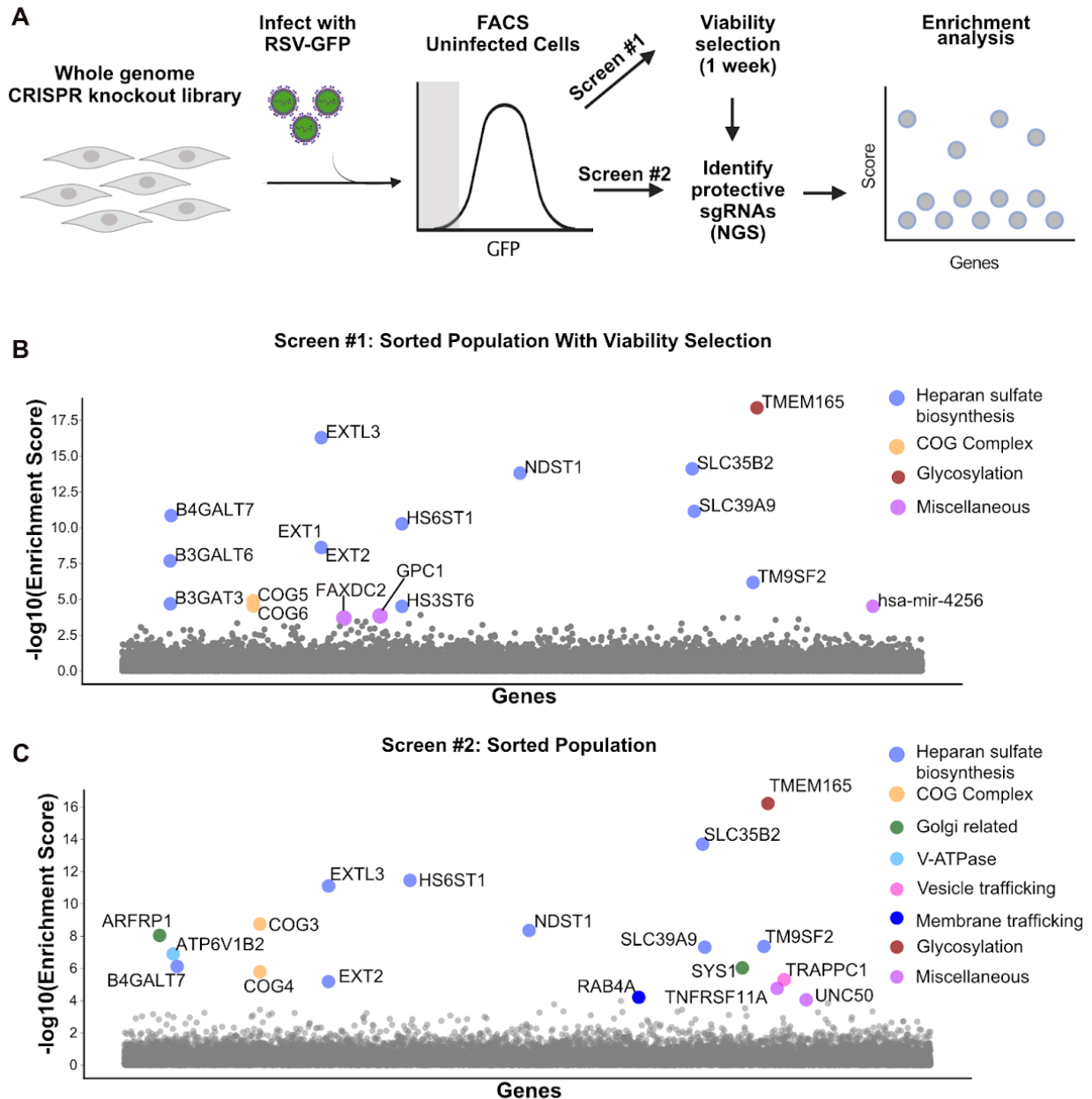
Figure 7.3: Genome-wide CRISPR/Cas9 genetic knockout screen in human A549 cells for identification of proviral host factors. A. Schematic of CRISPR/Cas9 screen; **B-C**. Gene enrichment analysis for Screen # 1 (sorted population with viability selection) and Screen # 2 (cells were harvested directly after sorting). Cells are colored by biological process/cellular location. The top hit for both screens, TMEM165, is highlighted in brown.
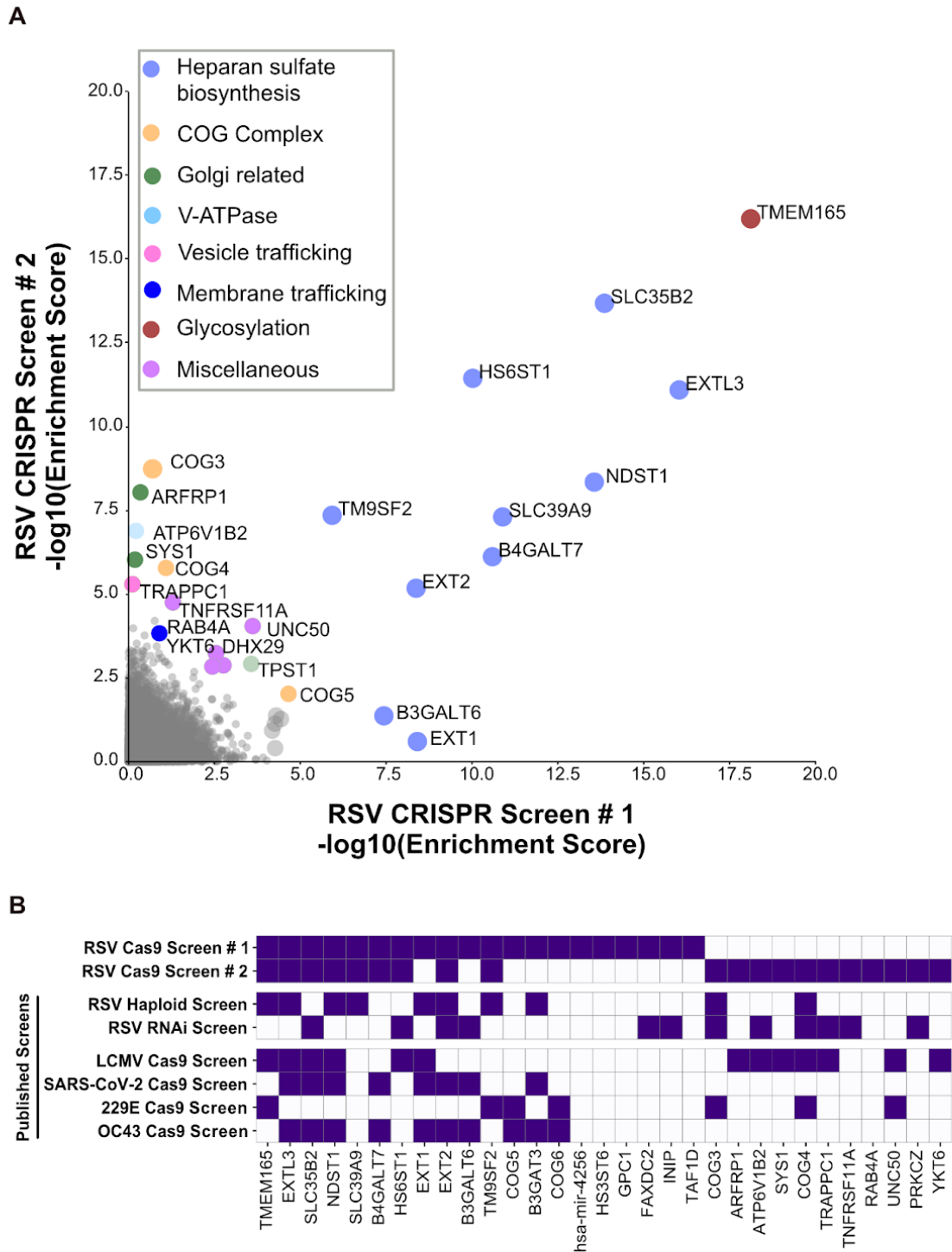
Figure 7.4: RSV Screen comparisons. A. Scatter plot comparing the results of Screen 1 and Screen 2. The -log10 MAGeCK enrichment score for each screen is plotted. **B**. Comparison of

the top hits from RSV CRISPR/Cas9 screens to previously published RSV haploid [17] and RNAi screens [16], as well as CRISPR genetic screens for LCMV, SARS-CoV-2, 229E and OC43 [23,24]. Each gene was evaluated for enrichment in RSV and other viral screens, and the color purple indicates the presence within each screen.
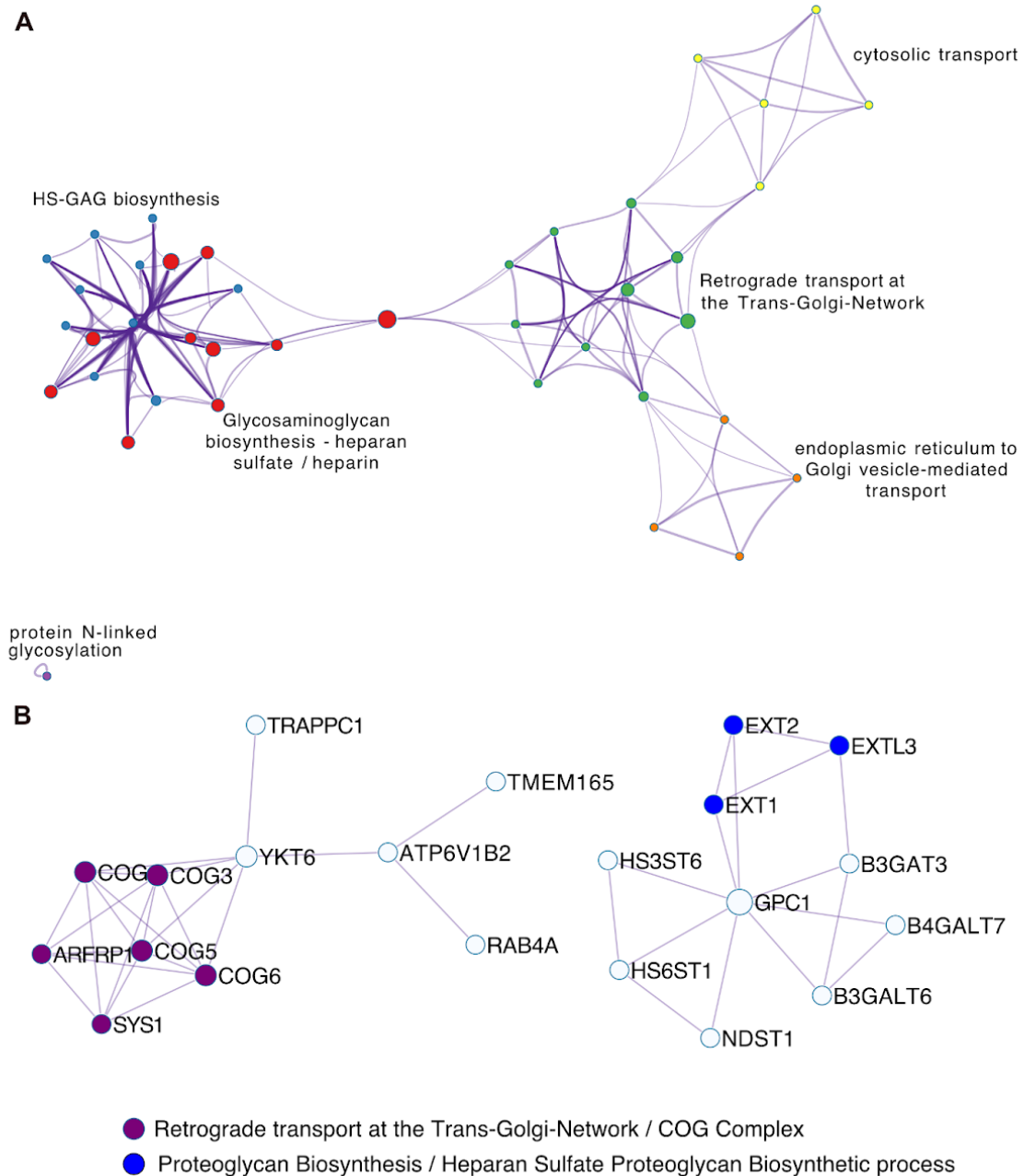
Figure 7.4: Metascape pathway and protein-protein interaction networks. A. Significantly enriched pathway clusters using Pathway and Process Enrichment Analysis (Metascape). Each circle represents a different pathway/process enrichment cluster. **B.** Protein-protein interaction

enrichment analysis (Metascape) reveals two clusters of networks, with each node representing a different host factor.

Table 7.1: Pearson's correlation with RSV NS1 in RSV infected cells.

Host gene	Pearson's r	p_val	p_val_adj
HERPUD1	0.46375971	4.34E-97	7.34E-93
HSPA5	0.39095692	4.26E-67	7.21E-63
MANF	0.3581509	7.62E-56	1.29E-51
SDF2L1	0.32867945	8.34E-47	1.41E-42
HSPA1A	0.31419291	1.03E-42	1.75E-38
DDIT3	0.30764286	6.18E-41	1.05E-36
DNAJB9	0.30162635	2.42E-39	4.10E-35
HSPE1	0.30140799	2.76E-39	4.67E-35

Table 7.2: Pearson's correlation with ISG15 in bystander cells.

Host gene	Pearson's r	p_val	p_val_adj
ISG15	1	0	0
MX1	0.75728932	0	0
IFIT1	0.74957469	0	0
IFI6	0.83264065	0	0
OAS1	0.72837121	1.0573e-320	2.38156934e-316
B2M	0.7076672	7.79E-295	1.75E-290

Host gene	Pearson's r	p_val	p_val_adj
IFI35	0.67444419	9.16E-258	2.06E-253
IFIT3	0.6728961	3.72E-256	8.39E-252
PLSCR1	0.65871691	7.30E-242	1.64E-237
IRF7	0.65747378	1.20E-240	2.70E-236
OASL	0.64631661	5.48E-230	1.23E-225
USP18	0.63992213	4.46E-224	1.00E-219
SP110	0.60596078	6.47E-195	1.46E-190
STAT1	0.60326795	9.41E-193	2.12E-188
UBE2L6	0.60044319	1.66E-190	3.74E-186
TAP1	0.58380912	1.02E-177	2.29E-173
HLA-C	0.58251939	9.29E-177	2.09E-172
SP100	0.57565582	1.02E-171	2.30E-167
EIF2AK2	0.57448826	7.16E-171	1.61E-166
PARP14	0.57436903	8.73E-171	1.97E-166
NT5C3A	0.57005606	1.08E-167	2.44E-163

Host gene	Pearson's r	p_val	p_val_adj
PSMB9	0.56978146	1.70E-167	3.82E-163
PSME2	0.56691698	1.81E-165	4.07E-161
IFI27	0.55268234	1.10E-155	2.48E-151
PNPT1	0.54918823	2.36E-153	5.31E-149
PARP9	0.54305024	2.51E-149	5.65E-145
OAS3	0.54267556	4.39E-149	9.90E-145
DDX58	0.54250017	5.71E-149	1.29E-144
PSMB8	0.53890336	1.19E-146	2.67E-142
HERC5	0.53610657	7.18E-145	1.62E-140
IFITM3	0.53443759	8.17E-144	1.84E-139

Methods

Cell Culture

HEp-2 (ATCC CCL-23) and A549 (ATCC CCL-185) cells were ordered from ATCC and cultured in DMEM supplemented with 10% FCS, penicillin, streptomycin, and glutamine. Cells were maintained at 37°C and 5% CO_2. Cells were mycoplasma negative using the MycoAlert Mycoplasma detection kit (Lonza).

RSV Propagation

For single-cell sequencing experiments, we used a genetically unmodified virus: Human respiratory syncytial virus (ATCC® VR-1540). For CRISPR/Cas9 screening experiments, we used Respiratory Syncytial Virus with EGFP (RSV-GFP5, ViraTree). Both isolates of RSV were amplified on HEp-2 cells, PEG precipitated, pelleted through a sucrose cushion during ultracentrifugation, resuspended and stored at -80°C. Viruses were titered on HEp-2 cells using the standard TCID50 assay. Both viral isolates were mycoplasma negative using the MycoAlert Mycoplasma detection kit (Lonza).

Single Cell RNA sequencing

A549 cells were infected with RSV at an MOI of 0.3 in serum-free media. After a two hour incubation, cells were washed with serum-free media, and fresh media with serum was replenished. To reduce batch effects, cells were sequentially infected and harvested at the same time point for single cell sequencing. Cells were trypsinized, murine NIH/3T3 (ATCC CRL-1658) were spiked in, and Multiseq sample barcoding was performed following McGinnis et al. protocol [26]. Cells were subsequently counted and the manufacturer's protocol for 10x Genomics 3' sequencing were performed. McGinnis et al. protocol was followed for Multiseq library

preparation. Samples were subsequently sequenced on the NovaSeq and aligned using 10x Genomics CellRanger v3.1.0 with default settings to hg38 reference genome concatenated with the RSV genome. Sample barcodes were demultiplexed following previously published protocols (McGinnis et al.). Downstream cell and gene filtering, log normalization, PCA and UMAP projection were performed using Scanpy v1.6.1 [27]. Differential expression was performed using MAST, and only included genes expressed in at least 10% of each population.

Genome-wide CRISPR/Cas9 Screen

A549 cells were lentivirally transduced with Cas9-BLAST (Addgene 52962; gift from Feng Zhang) and blasticidin selected. These A549-Cas9 lines were subsequently transduced with the human GeCKO v2 library (Addgene 1000000049; gift from Feng Zhang) and puromycin selected. This library was subsequently infected with RSV-EGFP. EGFP negative or low cells were sorted on the Sony SH800 cell sorter. For the first iteration of the screen with a viability selection, cells were cultured for 1 week and surviving cells were collected for downstream analysis. For the second screen, cells collected from the sorter were directly used for downstream analysis. We subsequently extracted genomic DNA, amplified guide sequences, performed library prep and sequenced on the Illumina NextSeq as described in Wang et al. [24].

Chapter 7 References

1. CDC. Learn about Respiratory Syncytial Virus Infection (RSV). 2020. Available at: https://www.cdc.gov/rsv/index.html. Accessed 14 September 2022.

2. Respiratory Syncytial Virus. **2021**; Available at: https://publications.aap.org/redbook/book/347/chapter/5755493/Respiratory-Syncytial-Virus. Accessed 14 September 2022.

3. Collins PL, Crowe JEJ. Respiratory syncytial virus and metapneumovirus. In: Fields virology. Philadelphia, PA.: Lippincott Williams & Wilkins, 2007.

4. Martínez I, Lombardía L, García-Barreno B, Domínguez O, Melero JAY 2007. Distinct gene subsets are induced at different time points after human respiratory syncytial virus infection of A549 cells. J Gen Virol 88:570–581.

5. Ioannidis I, McNally B, Willette M, et al. Plasticity and Virus Specificity of the Airway Epithelial Cell Immune Response during Respiratory Virus Infection. J Virol **2012**; 86:5422–5436.

6. Villenave R, Broadbent L, Douglas I, et al. Induction and Antagonism of Antiviral Responses in Respiratory Syncytial Virus-Infected Pediatric Airway Epithelium. J Virol **2015**; 89:12309–12318.

7. Hillyer P, Shepard R, Uehling M, et al. Differential Responses by Human Respiratory Epithelial Cell Lines to Respiratory Syncytial Virus Reflect Distinct Patterns of Infection Control. J Virol **2018**; 92:e02202-17.

8. Sun Y, López CB. The innate immune response to RSV: Advances in our understanding of critical viral and host factors. Vaccine **2017**; 35:481–488.

9. Spann KM, Tran KC, Collins PL. Effects of Nonstructural Proteins NS1 and NS2 of Human Respiratory Syncytial Virus on Interferon Regulatory Factor 3, NF-κB, and Proinflammatory Cytokines. J Virol **2005**; 79:5353–5362.

10. Spann KM, Tran K-C, Chi B, Rabin RL, Collins PL. Suppression of the Induction of Alpha, Beta, and Gamma Interferons by the NS1 and NS2 Proteins of Human Respiratory Syncytial Virus in Human Epithelial Cells and Macrophages. J Virol **2004**; 78:4363–4369.

11. Russell AB, Trapnell C, Bloom JD. Extreme heterogeneity of influenza virus infection in single cells. eLife **2018**; 7:e32303.

12. Zanini F, Pu S-Y, Bekerman E, Einav S, Quake SR. Single-cell transcriptional dynamics of flavivirus infection. eLife **2018**; 7:e32942.

13. Wyler E, Franke V, Menegatti J, et al. Single-cell RNA-sequencing of herpes simplex virus 1-infected cells connects NRF2 activation to an antiviral program. Nat Commun **2019**; 10:4878.

14. Hein MY, Weissman JS. Functional single-cell genomics of human cytomegalovirus infection. Nat Biotechnol **2022**; 40:391–401.

15. Wyler E, Mösbauer K, Franke V, et al. Transcriptomic profiling of SARS-CoV-2 infected human cell lines identifies HSP90 as target for COVID-19 therapy. iScience **2021**; 24:102151.

16. Anderson DE, Pfeffermann K, Kim SY, et al. Comparative Loss-of-Function Screens Reveal ABCE1 as an Essential Cellular Host Factor for Efficient Translation of Paramyxoviridae and Pneumoviridae. mBio **2019**; 10:e00826-19.

17. Hoffmann H-H, Schneider WM, Blomen VA, et al. Diverse viruses require the calcium transporter SPCA1 for maturation and spread. Cell Host Microbe **2017**; 22:460-470.e5.

18. Sunshine S, Puschnik AS, Replogle JM, et al. Systematic functional interrogation of SARS-CoV-2 host factors using Perturb-seq. 2022; :2022.07.15.500120. Available at: https://www.biorxiv.org/content/10.1101/2022.07.15.500120v1. Accessed 14 September 2022.

19. Lenschow DJ, Lai C, Frias-Staheli N, et al. IFN-stimulated gene 15 functions as a critical antiviral molecule against influenza, herpes, and Sindbis viruses. Proc Natl Acad Sci **2007**; 104:1371–1376.

20. Finak G, McDavid A, Yajima M, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol **2015**; 16:278.

21. Li W, Xu H, Xiao T, et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol **2014**; 15:554.

22. Thul PJ, Åkesson L, Wiking M, et al. A subcellular map of the human proteome. Science **2017**; 356:eaal3321.

23. Liu J, Knopp KA, Rackaityte E, et al. Genome-Wide Knockout Screen Identifies Human Sialomucin CD164 as an Essential Entry Factor for Lymphocytic Choriomeningitis Virus. mBio 13:e00205-22.

24. Wang R, Simoneau CR, Kulsuptrakul J, et al. Genetic Screens Identify Host Factors for SARS-CoV-2 and Common Cold Coronaviruses. Cell **2021**; 184:106-119.e14.

25. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun **2019**; 10:1523.

26. McGinnis CS, Patterson DM, Winkler J, et al. MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. Nat Methods **2019**; 16:619–626.

27. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol **2018**; 19:15.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

Sara Sunshine
—911900D4E13249A...   Author Signature

10/14/2022
Date

140