

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Understanding RNA viruses through functional genomics and next-generation sequencing techniques

Permalink

<https://escholarship.org/uc/item/1tj3b4qv>

Author

Liu, Chieh Ming Jamin

Publication Date

2022

Peer reviewed|Thesis/dissertation

Understanding RNA viruses through functional genomics and next-generation sequencing techniques

by

Chieh Ming Jamin Liu

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of

DOCTOR OF PHILOSOPHY

in

Bioengineering

in the

GRADUATE DIVISION

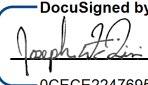
of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

AND


UNIVERSITY OF CALIFORNIA, BERKELEY

Approved:

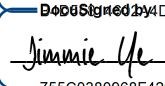
DocuSigned by:

0CECE2247695423...

Joseph DeRisi

Chair

DocuSigned by:


Laurent Coscoy

DocuSigned by:

755C0380968F429...

Jimmie Ye

Committee Members

Copyright 2022
by
Chieh Ming Jamin Liu

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Joseph DeRisi, for his support and mentorship through my graduate career. Joe has this infectious enthusiasm for science and discovery, exploration of the unknown, and pushing the boundaries of what is considered “technically possible.” He has built an extensive network of talented and driven researchers with diverse expertise, many of whom I have had the pleasure of collaborating. Thank you for creating such a unique and vibrant learning environment. My graduate school experience was vastly enriched through interacting with all these amazing scientists.

I must also thank my thesis committee members, Drs. Laurent Coscoy and Jimmie Ye, for their mentorship, support, and guidance on both the scientific and career-related aspects of my PhD. Your scientific expertise in virology and sequencing data analysis provided the necessary fresh perspectives to my projects; your continual encouragements allowed me to develop the confidence necessary for me to succeed. I would like to also express my sincere gratitude to my qualifying exam committee members, Drs. Britt Glaunsinger, Hana El-Samad, Aaron Streets, and committee chair, Dr. Dorian Liepmann. Thank you for helping me develop and refine my thinking as both an engineer and a scientist.

Science is never done alone, and without the help of my co-authors and collaborators, this dissertation would have never been completed. I would like to thank Dr. Peng Xu and Makiko Hatori for their help and expertise on microfluidic devices. I would also like to thank Dr. Diane Havlir. Dr. Genay Pilarowski, Dr. Douglas Black, Dr. Jeffrey Whitman, James Peng, Joslin Payan, Armando Mis, and the entire Unidos en Salud team for their invaluable contributions to helping us understand the spread of SARS-CoV-2 in San Francisco.

I would also like to thank the members of the DeRisi Lab for their collective support and wisdom. Dr. Kristeene Knopp, no words can express my gratitude for your decision to mentor

me and train me into a virologist. I am so thankful for supportive environment you have created in Team Virus. Thank you to Sara Sunshine, Dr. Matthew Laurie, and Dr. Hanna Retallack. There is no one else with whom I would rather experience a global pandemic. To the technicians, Sabrina Mann, Gustavo Reyes, Anthea Mitchell, Matthias Hapte-Selassie, and Grace Wang, it was my honor and pleasure to have worked with you and I look forward to hearing about your future endeavors. To my fellow graduate students, Andrew Kung, Dr. Sara Vazquez, and Robert Puccinelli, and my former rotation mentor, Dr. Valentina Gracia, I am thankful that we had each other while we embarked on this challenging training period. I would also like to thank the postdocs of the lab, Dr. Wesley Wu, Dr. Madhura Raghavan, Dr. Elze Rackaityte, and Dr. Sharline Madera, for their encouragements and guidance. And, of course, thank you to Manny De Vera and Brittany Worth, without whom I would not have had the resources to complete my research.

Mentorship can also come from outside of the home lab or institution, and I was lucky enough to have received guidance from so many scientists outside the DeRisi lab. Thank you to Dr. Andreas Puschnik, Dr. Manu Leonetti, Dr. Gorica Margulis, Dr. Michael Wilson, Dr. Jeremy Reiter, Dr. Eric Chow, Dr. James Nuñez, and Dr. Chun Yang. Your perspectives were ever helpful in keeping me grounded and your encouragements kept me motivated on the toughest of days.

To my wonderful cohort and friends from the UC Berkeley – UCSF Bioengineering Graduate Program, you were all a blast to hang out with. Thank you to Joao Ascensao, Nadia Ayad, Barry Fung, Dan Friedman, Jessie Liu, Connor Tsuchida, and Albert You for being such amazing and supporting friends. And I could never forget the administrative assistants and staff running the program behind the scenes. Thank you to Kristin Olson, SarahJane Taylor, Victoria Ross, Lisa Cabahug, and Rocio Sanchez for answering all my questions and making sure that I am slowly but surely progressing towards degree completion.

I also would like to thank my friends outside the program. To Amy Fan and Elisabeth Meyer: we have come a long way from Boston and being research technicians together. I am so glad to have shared this adventure together in the Bay Area. I would also like to thank David Bian, Valentin Hernandez, Eugene Kang, Elaine Kong, Harold Li, Sandy Lu, and Kevin Ng: Thank you for just believing in me and listening to me talk about my projects repeatedly and feign interest even when you were bored out of your minds.

There are countless others to thank: all my co-authors and collaborators, my friends and contemporaries, and my teachers and mentors from every stage of my life. You have all helped shape me into the scientist and engineer I am today. Thank you.

Of course, I completed my PhD without my family. To my mother, Meng-lan Hsu, and my father, Chun-chang Liu: thank you for your unwavering love and support. Thank you, Dr. Yi Ariel Liu, for being my favorite sister and becoming a medical doctor so I didn't have to. To my American family that took me under their wing in high school and never stopped, Mr. and Mrs. Meyers, Pog and Al, Edward and Lucy, and all your dogs and cats, goats, donkeys, horses and chickens: Thank you for supporting me always. And, to my in-laws, Na Chen and Yiwei Li, thank you for opening your home to me.

Finally, I would like to thank my husband, Dr. Ran Li, for his unwavering support. I do not listen to your advice enough. Thank you. I dedicate this work to you.

CONTRIBUTIONS

Chapter 2 of this dissertation is a reprint of the material largely as it appears in:

Liu J, Knopp KA, Rackaityte E, Wang CY, Laurie MT, Sunshine S, Puschnik AS, and DeRisi JL.

Genome-wide knockout screen identifies human sialomucin CD164 as an essential entry factor for lymphocytic choriomeningitis virus. *mBio* 2022; May 3:e0020522. doi:

10.1128/mbio.00205-22. Epub ahead of print.

Chapter 3 includes contributions by Dr. Kristeene Knopp and Dr. Joseph DeRisi.

Chapter 4 of this dissertation is a reprint of the material as it appears 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, The CLIAHUB

Consortium. Performance Characteristics of a Rapid SARS-CoV-2 Antigen Detection

Assay at a Public Plaza Testing Site in San Francisco. *J Infect Dis* 2021; 223.7: 1139-

1144. doi:10.1093/infdis/jiaa802.

Chapter 5 of this dissertation is a reprint of the material as it appears 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 JL, and 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; 74:32–39.

doi:10.1093/cid/ciab283. *Authors contributed equally

Chapter 6 of this dissertation is a reprint of the material as it appears 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 Jr. 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; jjab635. doi:10.1093/infdis/jiab635.

Chapter 7 of this dissertation is a reprint of the material as it appears 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 MA, Havlir D, and DeRisi JL. Severe acute respiratory syndrome

coronavirus 2 (SARS-CoV-2) transmission dynamics and immune responses in a

household of vaccinated persons. *Clin Infect Dis* 2022; ciac029. doi:10.1093/cid/ciac029.

Chapter 8 of this dissertation is reprinted largely as it appears in:

Garcia VE*, Liu J*, DeRisi JL. Low-Cost Touchscreen Driven Programmable Dual Syringe

Pump for Life Science Applications. *HardwareX* 2018; 4: e00027.

doi:10.1093/ofid/ofx162.082. *Authors contributed equally

*If I have seen a little further
it is by standing on the shoulders of Giants.*
- Isaac Newton

Understanding RNA viruses through functional genomics and next-generation sequencing techniques

Chieh Ming Jamin Liu

Abstract

Viruses are obligate intracellular pathogens that must enter and hijack host cells to survive. Many are human pathogens and pose a direct threat to human health worldwide, as seen with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Currently, viral diagnoses are slow, mostly made by ruling out bacterial infections, and treatment methods consists mainly of symptom management. By applying next-generation sequencing (NGS) techniques to virology, we hope to better comprehend how viral infections occur, thereby elucidating new therapeutic targets, as well to diagnose, track, and trace emerging viral variants to guide public health policies. Here, we describe the application of sequencing techniques to two different viruses: lymphocytic choriomeningitis virus (LCMV) and SARS-CoV-2.

We begin with a whole-genome functional screen to identify the host repertoire involved in LCMV infection in chapter 2. This high-throughput technique allowed us to rapidly identify known and novel host factors, including sialomucin CD164. We followed up on this protein to highlight its critical role in LCMV entry and speculate on its role in congenital infections. We then developed the foundations necessary for a novel screening modality to elucidate the host factors involved in viral assembly and budding (chapter 3). By combining droplet-based microfluidics and whole-genome screening techniques, we demonstrate the modules necessary for exploring this previously understudied stage of the viral life cycle.

The contents of chapters 3 through 7 describes the collaborative work completed during the early days of the SARS-CoV-2 global pandemic with community leaders and UCSF

clinicians through the Unidos en Salud (UeS) organization. Prior to the establishment of publicly available, accessible, and rapid testing facilities, we evaluated the performance of the Abbott BinaxNOW rapid antigen tests for coronavirus disease 19 (COVID-19) to detect the virus among symptomatic and asymptomatic individuals by comparing the results of the rapid antigen test with that of a reverse-transcription polymerase chain reaction (RT-PCR) test (Chapter 4). As the pandemic persisted, emerging variants of SARS-CoV-2 were identified by viral genome sequencing technologies. Within our local community, we identified a novel “West Coast” variant, now known as B.1.427 and B.1.429 or the epsilon variant (Chapter 5). Variant distributions combined the extensive metadata collected by UeS revealed an increase in relative household attack rates consistent with a modest transmissibility increase of the epsilon variant.

As vaccines and boosters became publicly available, novel variants of concern (VOC) also emerged. In chapter 6, we quantified the efficacy of mRNA vaccines and prior exposures against circulating VOC using an antibody neutralization assay. While vaccines prevented severe disease effectively, post-vaccination breakthrough COVID-19 infections remain a public health concern. We present a detailed case study describing the transmission of the gamma variant (P.1) within an immunized family over the course of several weeks. This characterization revealed not only the complexity of transmission dynamics, but also the necessity of understanding relevant co-morbidities, including auto-immunity to type-1 interferon (Chapter 7).

Finally, in chapter 8, we return to our engineering roots and describe an instrument development project, where we designed and assembled a low-cost syringe pump customized specifically to automate production of translationally active cell lysates.

TABLE OF CONTENTS

1 INTRODUCTION	1
1.1 VIROLOGY IN THE ERA OF HIGH-THROUGHPUT SEQUEUNCING	1
1.2 LYMPHOCYTIC CHORIOMENINGITIS VIRUS	2
1.3 SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS 2.....	4
1.4 SUMMARY OF WORK AND FINDINGS	5
1.4 REFERENCES.....	7
2 GENOME-WIDE KNOCKOUT SCREEN IDENTIFIES HUMAN SIALOMUCIN CD164 AS AN ESSENTIAL ENTRY FACTOR FOR LYMPHOCYTIC CHORIOMENINGITIS VIRUS	11
2.1 ABSTRACT	11
2.2 IMPORTANCE	12
2.3 INTRODUCTION.....	12
2.4 RESULTS.....	15
<i>2.4.1 CRISPR KO screens identify host factors for LCMV infection</i>	<i>15</i>
<i>2.4.2 Pseudotyped viral infection shows that CD164 is a LCMV-specific mammarenavirus human entry factor</i>	<i>18</i>
<i>2.4.3 N-linked glycosylation within the cysteine-rich domain is critical for LCMV infection.....</i>	<i>20</i>
<i>2.4.4 CD164 is highly expressed in human placenta and mediates LCMV infection in placental cells</i>	<i>21</i>
2.5 DISCUSSION.....	22
2.6 MATERIALS AND METHODS.....	25
<i>2.6.1 Cell lines</i>	<i>25</i>

2.6.2 Genome-wide CRISPR screen	26
2.6.3 Generation of monoclonal KO cell lines.....	27
2.6.4 Plasmids, cloning, and lentivirus production	27
2.6.5 Compound inhibition and antibody neutralization	28
2.6.6 Generation of Arenavirus pseudotyped vesicular stomatitis virus	28
2.6.7 Flow cytometry analysis of viral infection assays.....	29
2.6.8 Indirect Immunofluorescence of Human Placenta	30
2.6.9 Western Blots.....	30
2.6.10 Quantification and Statistical Analysis	30
2.7 ACKNOWLEDGEMENTS	31
2.8 FIGURES	32
2.9 TABLES	45
2.10 REFERENCES.....	54
3 IDENTIFICATION OF HOST FACTORS INVOLVED IN LCMV VIRAL EGRESS AND FUTURE DIRECTIONS	62
3.1 BACKGROUND	62
3.2 EXPERIMENTAL APPROACHES.....	64
3.2.1 Engineered split reporter system allows for identification of successful viral egress. 64	
3.2.2 Droplet-based microfluidics allows independent assays to be conducted in a high- throughput manner.....	66
3.2.3 Novel droplet-based technologies allow for efficient sorting of double emulsions.....	72
3.3 IDENTIFICATION OF NOVEL HOST FACTORS INVOLVED IN VIRAL EGRESS	73
3.3.1 Droplet-based high-throughput whole-genome function screen for viral egress factors	73

3.3.2	<i>Candidate-based high-throughput function screen for viral egress factors.....</i>	75
3.3.3	<i>Expected results</i>	76
3.4	CURRENT CHALLENGES, IMPACT, AND FUTURE DIRECTIONS	77
3.5	REFERENCES.....	78
4	PERFORMANCE CHARACTERISTICS OF A RAPID SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS 2 ANTIGEN DETECTION ASSAY AT A PUBLIC PLAZA TESTING SITE IN SAN FRANCISCO.....	82
4.1	ABSTRACT	82
4.2	INTRODUCTION.....	83
4.3	METHODS	84
4.3.1	<i>Study population and specimen collection.....</i>	<i>84</i>
4.3.2	<i>Laboratory testing for SARS-CoV-2.....</i>	<i>84</i>
4.3.3	<i>Field testing using Binax-CoV2 Assay</i>	<i>84</i>
4.3.4	<i>Titration of in vitro cultured SARS-CoV-2 on Binax-CoV2 Cards.....</i>	<i>85</i>
4.3.5	<i>N Protein Titration Assay</i>	<i>85</i>
4.3.6	<i>Ethics Statement.....</i>	<i>85</i>
4.4	RESULTS.....	86
4.4.1	<i>Binax-CoV2 Performance Using a Titration of in vitro cultured SARS-CoV-2</i>	<i>86</i>
4.4.2	<i>Community RT-PCR testing results</i>	<i>86</i>
4.4.3	<i>Comparison of RT-PCR and Binax-CoV2 testing results from community testing</i>	<i>86</i>
4.4.4	<i>Sensitivity and specificity</i>	<i>88</i>
4.4.5	<i>Evaluation of Binax-CoV2 lot-to-lot variation</i>	<i>88</i>
4.5	DISCUSSION.....	89
4.6	ACKNOWLEDGEMENTS	90

4.7 FIGURES	92
4.8 REFERENCES.....	95
5 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	97
5.1 ABSTRACT	97
5.2 INTRODUCTION.....	98
5.3 METHODS	100
5.3.1 <i>Study setting and population</i>	100
5.3.2 <i>SARS-CoV-2 genomic sequence recovery and consensus genome generation</i>	100
5.3.3 <i>Household attack rate analyses</i>	101
5.3.4 <i>Bayesian Phylogenetic Analysis</i>	101
5.3.5 <i>Ethics statement</i>	102
5.4 RESULTS.....	102
5.4.1 <i>Low-Barrier SARS-CoV-2 Testing and Sequencing</i>	102
5.4.2 <i>Disease Severity</i>	104
5.4.3 <i>Household Secondary Attack Rate</i>	104
5.4.4 <i>Estimation of Reproductive Number</i>	106
5.5 DISCUSSION.....	106
5.6 ACKNOWLEDGEMENTS	109
5.7 SUPPLEMENTARY METHODS.....	110
5.7.1 <i>SARS-CoV-2 genomic sequence recovery</i>	110
5.7.2 <i>SARS-CoV-2 consensus genome generation</i>	110
5.7.3 <i>Bayesian Phylogenetic Analysis</i>	111

5.8 FIGURES	113
5.9 TABLES	117
5.10 REFERENCES.....	125
6 SARS-COV-2 VARIANT EXPOSURES ELICIT ANTIBODY RESPONSES WITH DIFFERENTIAL CROSS-NEUTRALIZATION OF ESTABLISHED AND EMERGING STRAINS INCLUDING DELTA ANDOMICRON	130
6.1 ABSTRACT	130
6.2 INTRODUCTION.....	131
6.3 METHODS	132
<i>6.3.1 Serum collection</i>	<i>132</i>
<i>6.3.2 Pseudovirus production</i>	<i>133</i>
<i>6.3.2 Pseudovirus neutralization experiments</i>	<i>134</i>
<i>6.3.2 Data analysis</i>	<i>134</i>
6.4 RESULTS.....	135
6.5 DISCUSSION.....	136
6.6 ACKNOWLEDGEMENTS	138
6.7 FIGURES	140
6.8 TABLES	142
6.9 REFERENCES.....	147
7 SARS-COV-2 TRANSMISSION DYNAMICS AND IMMUNE RESPONSES IN A HOUSEHOLD OF VACCINATED PERSONS.....	149
7.1 ABSTRACT	149
7.2 INTRODUCTION.....	149
7.3 DESCRIPTION OF INDIVIDUALS IN THE STUDY HOUSEHOLD	150

7.4 RESULTS.....	152
7.5 DISCUSSION.....	153
7.6 ACKNOWLEDGEMENTS	155
7.7 SUPPLEMENTARY METHODS.....	156
7.7.1 <i>Specimen collection and processing</i>	156
7.7.2 <i>Whole viral genome sequencing</i>	157
7.7.3 <i>Pseudovirus neutralization assay</i>	158
7.7.4 <i>Anti-Type I Interferon Alpha 2 (anti-IFN-α2) Radioligand Binding Assay (RLBA)</i>	158
7.8 FIGURES	159
7.9 TABLES	162
7.10 REFERENCES.....	165
8 LOW-COST TOUCHSCREEN DRIVEN PROGRAMMABLE DUAL SYRINGE PUMP FOR LIFE SCIENCE APPLICATIONS	168
8.1 ABSTRACT.....	168
8.2 HARDWARE IN CONTEXT.....	169
8.3 HARDWARE DESCRIPTION.....	170
8.4 DESIGN FILES.....	171
8.5 BILL OF MATERIALS.....	172
8.6 BUILD INSTRUCTIONS.....	172
8.6.1 <i>Preparation of electronic components</i>	173
8.6.2 <i>Hardware preparation</i>	173
8.6.3 <i>Hardware assembly</i>	174
8.6.4 <i>Software setup</i>	176
8.7 OPERATION INSTRUCTIONS	177

8.7.1 Operation instructions for the provided GUI.....	177
8.7.2 Customization suggestions for operation.....	178
8.8 VALIDATION AND CHARACTERIZATION	179
8.8.1 Characterization of the PDSP as a syringe pump.....	179
8.8.2 Validation of the PDSP as a cell homogenizer	180
8.8.3 Conclusion and device overview.....	181
8.9 ACKNOWLEDGEMENTS AND FUNDING.....	181
8.10 FIGURES	182
8.11 TABLES	188
8.12 REFERENCES.....	192
9 APPENDIX I: DATA AVAILABILITY	194

LIST OF FIGURES

Figure 2.1 Genome-wide CRISPR loss-of-function screen in human cells identify host factors important for LCMV infection.....	33
Figure 2.2 Infection of KO cell lines with a panel of mammarenavirus GP pseudotyped virus ..	34
Figure 2.3 CD164 functional region determination through antibody binding, domain deletion, and alanine mutagenesis.....	35
Figure 2.4 Characterization of CD164 as a therapeutic target in human placenta.....	37
Figure 2.S1 Validation of recombinant virus rLCMV-mCherry infectivity.....	38
Figure 2.S2 Additional hit validation and characterization of gene-edited cells.....	40
Figure 2.S3 Characterization of $\Delta CD164$, $\Delta DAG1$, and $\Delta CD164/\Delta DAG1$ double KO cells.....	41
Figure 2.S4 Characterization of CD164 domain deletion and alanine mutagenesis add backs.....	42
Figure 2.S5 Characterization of CD164 in placenta tissue and cell line.....	43
Figure 2.S6 Gene enrichment score comparison between two whole-genome CRISPR screens for LCMV host factors.....	44
Figure 4.1 Titration of in vitro grown severe acute respiratory syndrome coronavirus 2 and detection with Binax-CoV2 assay.....	92
Figure 4.2 Comparison of Binax-CoV2 test with quantitative reverse-transcription polymerase chain reaction (RT-PCR) test	93
Figure 4.S1 Variability of signal intensity in Binax-CoV2 card lots	94
Figure 5.1 Testing catchment area.....	113
Figure 5.2 Variants observed at 24th & Mission.....	114
Figure 5.S1 NextClade Phylogeny	115
Figure 5.S2 Viral load by Ct value	116

Figure 6.1 Neutralization of D614G and B.1.429 pseudoviruses by serum from individuals
with different exposures 140

Figure 6.2 Change in variant pseudovirus neutralization titer relative to D614G 141

Figure 7.1 Serum samples from household individuals reveal diverse neutralization
capabilities as well as presence of anti-IFN- α 2 auto-antibodies in Individual 1 160

Figure 7.S1 Sequencing depth and coverage of recovered SARS-CoV-2 genomes. 161

Figure 8.1 Assembled PDSP 182

Figure 8.2 3D render of all 3D printed PDSP components..... 183

Figure 8.3 Electrical wiring schematic of the PDSP 184

Figure 8.4 Key PDSP assembly steps as described in 8.6.3 Hardware assembly..... 185

Figure 8.5 PDSP GUI 186

Figure 8.6 PDSP characterization for lysate generation..... 187

LIST OF TABLES

Table 2.S1 rLCMV-mCherry whole genome CRISPR screen results	45
Table 2.S2 Gene ontology overrepresentation analysis.....	47
Table 2.S3 Antibodies used in this study	49
Table 2.S4 Oligonucleotides used in this study.....	50
Table 5.1 Characteristics of Households Included in the Household Attack Rate Analysis, Stratified by Strain.....	117
Table 5.2 Secondary Household Attack Rates for West Coast Variants, Combined and Disaggregated by B.1.427 and B.1.429	118
Table 5.3 Secondary Attack Rate Disaggregated by Covariates	119
Table 5.S1 Characteristics of all persons (N=8,822; 10.0% with positive result) tested at the community-based testing site between January 10-29	120
Table 5.S2 Sample sequencing summary and viral cycle threshold characteristics	121
Table 5.S3 Amino acid mutations observed in the spike gene and count of sequences per mutation in each study	122
Table 5.S4 Individual characteristics of all persons tested (both positive and negative, including index case) living in one of the 319 households meeting inclusion criteria for household secondary attack rate analyses, stratified by strain classification of the household	123
Table 5.S5 Adjusted attack rates from sensitivity analysis using Targeted Maximum Likelihood and Super Learning	124
Table 6.S1 Serum sample metadata and exposure information	142
Table 6.S2 Spike gene mutations for each SARS-CoV-2 pseudovirus.....	145

Table 6.S3 Table of measured pseudovirus neutralization titers in each serum group (mean of three replicates)	146
Table 7.S1 Study household description and metadata.....	162
Table 7.S2 Characteristic mutations of the Gamma variant in SARS-CoV-2 positive individuals.	163
Table 7.S3 Healthy vaccinated cohort metadata.....	164
Table 8.1 Design files for all 3D printed parts	188
Table 8.2 List of all purchased items with number required and cost	189
Table 8.3 Minimum and maximum theoretical flow rates by the PDSP based on syringe size	191

LIST OF ABBREVIATIONS

ARDS	Acute respiratory distress syndrome
BINAX-CoV2	Abbott BinaxNOW COVID-19 Ag Card
BinaxNOW	Abbott BinaxNOW COVID-19 Ag Card
COVID-19	Coronavirus disease 19
CRISPR	clustered regularly interspaced short palindromic repeats
GO	Gene ontology
GPC (LCMV)	Glycoprotein complex
GTOV	Guanarito virus
GUI	Graphical user interface
IFN	Interferon
KO	Knockout
LASV	Lassa virus
LCMV	Lymphocytic choriomeningitis virus
mAb	Monoclonal antibody
MACV	Machupo virus
NP	Nucleoprotein
S (SARS-CoV-2)	Spike protein
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
VSV	Vesicular stomatitis virus

1 INTRODUCTION

1.1 VIROLOGY IN THE ERA OF HIGH-THROUGHPUT SEQUENCING

Rapid technological advancements in DNA sequencing since the completion of the human genome has transformed our ability to study biology and disease. High-throughput sequencing, or next-generation sequencing, describes the ability to sequence millions of DNA fragments in parallel and thus produce an enormous amount of data at once [1]. Major declines in reagent, computing, and other associated costs have allowed this technology to become ubiquitous in biomedical research settings. Production of genomic data has long outpaced Moore's Law and is on track to becoming the largest source of digital information in the world by 2025 [2]. The impact of high-throughput sequencing on all fields of biology cannot be understated. With the ability to simultaneously read out the results of parallelized genetic assays, large-scale experiments such as genome-wide screens and population sequencing became possible. In lieu of candidate-based approaches, whole-genome functional screens combined with the accessibility of genetic editing techniques like CRISPR have become the norm in molecular biology.

Next-generation sequencing have also transformed the way in which virology is studied. Now, all at once, using a whole-genome functional screen, a virologist can simultaneously assay which human genes play a role during viral infections [3–5]. In the case of an emerging virus, scientists can sequence hundreds of thousands of samples in parallel to observe rapid evolutionary changes in the population [6]. This powerful tool will certainly continue to advance discoveries in basic science that expand our understanding of the molecular mechanisms of infections and lead to novel translational therapies for viral disease.

In this dissertation, I present my work and efforts in applying high-throughput sequencing technologies to the study of two viruses in particular: lymphocytic choriomeningitis virus (LCMV), by using whole-genome functional screens, and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), by conducting viral genome sequencing to track emerging variants.

1.2 LYMPHOCYTIC CHORIOMENINGITIS VIRUS

LCMV is a member of the arenaviridae family and classified into the genera of *Mammarenavirus*, whose natural hosts are predominantly rodents [7,8]. Mammarenaviruses can be further divided into Old World (OW), which are mainly indigenous to Africa, and New World (NW), which are indigenous to the Americas, based on antigenic properties [9]. Known as the prototypic arenavirus, LCMV is a OW virus that can be found worldwide due to the global spread of its natural reservoir, the house mice or *Mus musculus* [10]. LCMV is a pleomorphic, enveloped, negative-stranded RNA virus. It has a bi-segmented ambi-sense genome that encodes only four genes: the viral RNA-dependent RNA polymerase (L) and a small RING finger protein (Z) that is functionally equivalent to the matrix protein found in many enveloped RNA viruses are found on the large (L) genome segment; and the viral nucleoprotein (NP) and the glycoprotein complex (GPC) which are found on the small (S) genome segment [11,12].

LCMV was first isolated in 1933 by Charles Armstrong when it was erroneously named as the causative agent of an epidemic encephalitis in St. Louis [13,14]. Other studies that made observations towards both the virus and the immune response in mice and humans soon followed in 1936 [10,15]. One such observation pertained to the differences causing persistent versus acute viral infections [16]. These discoveries gave way to LCMV attaining a leading role in modern immunology research. Studies using LCMV as a model led to further understanding

of immunological tolerance due to *in utero* infections as well as the importance of viral variance in antibody production levels [17]. Furthermore, isolation of the variant LCMV Clone 13 gave way to research that demonstrated that viral variance can also affect persistence as well as organ distribution. Modern sequencing techniques then revealed that LCMV Armstrong and Clone 13 differed by a mere two amino acid changes, a K→Q substitution in the L protein and a F→L substitution in the GP [18,19]. These two differences were critical in producing the biological phenotype of acute versus persistent infection and this concept, where single amino acid changes in a virus can dramatically alter the resulting pathogenicity has been extended to many other fields such as HIV [17].

While LCMV is well-studied in the context of immunological response, understanding of its viral reproduction process remains incomplete. Dystroglycan (DAG1), a widely expressed cell adhesion molecule, is recognized as the main attachment factor for viral entry by LCMV, Lassa Virus (LASV) and several other NW mammarenaviruses [7]. However, knockout of DAG1 was insufficient in preventing LCMV infection, thereby alluding to possible other attachment factors. During viral entry, viral and cell membranes fuse to allow viral ribonucleic proteins to be deposited into the cytoplasm. In this step, mammarenaviruses Lassa virus and Lujo virus requires the GP to bind to late endosomal resident proteins LAMP1 and CD36, respectively, in a low pH environment [4,20]. It is unknown whether LCMV also requires a receptor switch in the late endosome compartment for membrane fusion. Although many genes have been described to be relevant for LCMV infection (e.g., EIF4A1 plays a critical role in transcription and translation of viral genes, cellular site 1 protease/subtilisin kexin isozyme-1 (S1P/SKI-1) mediates cleavage of the GP complex for viral egress, and TSG101 from the host secretory pathway has an alternative function viral assembly [21–23]), many other host factors surely play a relevant role in this process but have yet to be described. As such, further studies are necessary for elucidating the complete life cycle of LCMV.

1.3 SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS 2

SARS-CoV-2 is an enveloped single-stranded RNA virus that is the causative agent of the coronavirus disease 19 (COVID-19) pandemic [24,25]. Following its emergence in human populations in December 2019 in Wuhan, China, it has spread globally to all populated continents. COVID-19 primarily manifests as a respiratory illness, infecting upper and lower airways and leading to viral pneumonia or acute respiratory distress syndrome (ARDS) [26,27]. As of May 2022, SARS-CoV-2 has infected over 529 million people worldwide leading to more than 6 million deaths [28].

The SARS-CoV-2 pandemic is a rapidly developing situation. Since the virus was first identified using next-generation sequencing techniques [29], over 11 million viral genomes have been deposited on the online database GISAID [30]. Additionally, the World Health Organization (WHO) has recognized 5 variants of concern (Alpha, Beta, Gamma, Delta, and Omicron) each of which are differ significantly from the original strain [31]. Such large-scale global monitoring of SARS-CoV-2 evolution has only been made possible due to the accessibility of high-throughput sequencing in public and research institutions.

Significant investments have been made to fast-track diagnostic, therapy, and vaccine development for SARS-CoV-2 [32,33], many proving fruitful in abating the spread of the viral infection. However, as novel variants continue to emerge and dissipate throughout the world, it is imperative that efforts to continue to track and evaluate SARS-CoV-2 strains do not stop. Scientists must continue assessing the efficacy of vaccines against new circulating variants of concern and evaluate the levels protective antibodies generated from prior exposures. Researchers must continue exploring different avenues for new therapies for not only symptom relief but also to treat the infection at its source. Only by working together, would we be able to quell this pandemic and protect our most vulnerable populations.

1.4 SUMMARY OF WORK AND FINDINGS

The use of next-generation sequencing techniques has been transformative in aiding our understanding of biology and the natural world. Here, I describe my work in leveraging this technology to study the field of virology. In chapters 2 and 3, I used genome-wide CRISPR knockout screens to determine the host repertoire involved in LCMV infection [34]. Additionally, I blueprinted a technology development project in an attempt to study a rarely assayed stage of the viral life cycle: viral assembly and budding. By combining CRISPR screening technologies with microfluidic droplet generation, I designed an assay able to specifically identify host factors that play a role in viral egress for LCMV.

In chapters 4 through 7, I made used next-generation sequencing to identify and track variants of concern emerging and spreading in a local community in San Francisco. In an ongoing collaboration with community members, clinicians, and UCSF researchers (Unidos en Salud), we provided low-barrier and accessible testing to members of the community using rapid antigen tests [35]. Through this partnership, we were able to assess the performance characteristics of the Abbott BinaxNOW COVID-19 Ag Cards as well as determine the attack rate of the epsilon variant as it peaked in the California [36]. We were also able to explore the antibody responses elicited by various SARS-CoV-2 variants and assess their protective effects in new and emerging strains [37]. Finally, using the extensive metadata collected by our community partners, we were able to present a detailed characterization of transmission and immunity among individuals in a vaccinated household [38].

The work described in this thesis is a culmination of the creativity and passion of the many talented scientists and clinicians who came together to work in the fast-paced, collaborative, and multi-disciplinary environment that is the DeRisi lab. It was truly my great pleasure to be a part of this exciting process to not only learn and develop as a bioengineer and virologist, but also to also be continuously exposed and challenged with new ideas and scientific

discoveries. Being part of a group studying viruses during a global pandemic truly showed me the importance of basic science research. I am immensely proud of the work we have accomplished together.

1.4 REFERENCES

1. Behjati S, Tarpey PS. What is next generation sequencing? Arch Dis Child Educ Pract Ed **2013**; 98:236–238.
2. Morey M, Fernández-Marmiesse A, Castiñeiras D, Fraga JM, Couce ML, Cocho JA. A glimpse into past, present, and future DNA sequencing. Mol Genet Metab **2013**; 110:3–24.
3. Savidis G, McDougall WM, Meraner P, et al. Identification of Zika Virus and Dengue Virus Dependency Factors using Functional Genomics. Cell Rep **2016**; 16:232–246.
4. Raaben M, Jae LT, Herbert AS, et al. NRP2 and CD63 Are Host Factors for Lujo Virus Cell Entry. Cell Host Microbe **2017**; 22:688-696.e5.
5. Staller E, Sheppard CM, Neasham PJ, et al. ANP32 Proteins Are Essential for Influenza Virus Replication in Human Cells. J Virol **2019**; 93:e00217-19.
6. Quick J. nCoV-2019 sequencing protocol v2 (GunIt). **2020**; Available at: <https://www.protocols.io/view/ncov-2019-sequencing-protocol-v2-bdp7i5rn>. Accessed 11 September 2021.
7. Radoshitzky SR, Buchmeier MJ, Charrel RN, et al. ICTV Virus Taxonomy Profile: Arenaviridae. J Gen Virol **2019**; 100:1200–1201.
8. Shi M, Lin X-D, Chen X, et al. The evolutionary history of vertebrate RNA viruses. Nature **2018**; 556:197–202.
9. Burri DJ, da Palma JR, Seidah NG, et al. Differential recognition of Old World and New World arenavirus envelope glycoproteins by subtilisin kexin isozyme 1 (SKI-1)/site 1 protease (S1P). J Virol **2013**; 87:6406–6414.
10. Traub E. A FILTERABLE VIRUS RECOVERED FROM WHITE MICE. Science **1935**; 81:298–299.

11. Meyer BJ, de la Torre JC, Southern PJ. Arenaviruses: genomic RNAs, transcription, and replication. *Curr Top Microbiol Immunol* **2002**; 262:139–157.
12. Knipe DM, Howley PM, editors. *Fields virology*. 6th ed. Philadelphia, PA: Wolters Kluwer/Lippincott Williams & Wilkins Health, 2013.
13. Luby JP. St. Louis encephalitis. *Epidemiol Rev* **1979**; 1:55–73.
14. Zhou X, Ramachandran S, Mann M, Popkin DL. Role of lymphocytic choriomeningitis virus (LCMV) in understanding viral immunology: past, present and future. *Viruses* **2012**; 4:2650–2669.
15. Rivers TM, Scott TF. MENINGITIS IN MAN CAUSED BY A FILTERABLE VIRUS : II. IDENTIFICATION OF THE ETIOLOGICAL AGENT. *J Exp Med* **1936**; 63:415–432.
16. Weigand H, Hotchin J. Studies of lymphocytic choriomeningitis in mice. II. A comparison of the immune status of newborn and adult mice surviving inoculation. *J Immunol* **1961**; 86:401–406.
17. Tishon A, Salmi A, Ahmed R, Oldstone MB. Role of viral strains and host genes in determining levels of immune complexes in a model system: implications for HIV infection. *AIDS Res Hum Retroviruses* **1991**; 7:963–969.
18. Kunz S, Sevilla N, Rojek JM, Oldstone MBA. Use of alternative receptors different than alpha-dystroglycan by selected isolates of lymphocytic choriomeningitis virus. *Virology* **2004**; 325:432–445.
19. Salvato M, Borrow P, Shimomaye E, Oldstone MB. Molecular basis of viral persistence: a single amino acid change in the glycoprotein of lymphocytic choriomeningitis virus is associated with suppression of the antiviral cytotoxic T-lymphocyte response and establishment of persistence. *J Virol* **1991**; 65:1863–1869.
20. Jae LT, Raaben M, Herbert AS, et al. Virus entry. Lassa virus entry requires a trigger-induced receptor switch. *Science* **2014**; 344:1506–1510.

21. Beyer WR, Pöpplau D, Garten W, von Laer D, Lenz O. Endoproteolytic processing of the lymphocytic choriomeningitis virus glycoprotein by the subtilase SKI-1/S1P. *J Virol* **2003**; 77:2866–2872.
22. Knopp KA, Ngo T, Gershon PD, Buchmeier MJ. Single nucleoprotein residue modulates arenavirus replication complex formation. *mBio* **2015**; 6:e00524-00515.
23. Pasqual G, Rojek JM, Masin M, Chatton J-Y, Kunz S. Old world arenaviruses enter the host cell via the multivesicular body and depend on the endosomal sorting complex required for transport. *PLoS Pathog* **2011**; 7:e1002232.
24. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**; 579:265–269.
25. Zhou P, Yang X-L, Wang X-G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**; 588:E6.
26. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. *JAMA* **2020**; 324:782–793.
27. Drosten C, Günther S, Preiser W, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* **2003**; 348:1967–1976.
28. Johns Hopkins University. COVID-19 Dashboard. **2022**; Available at: <https://coronavirus.jhu.edu/map.html>. Accessed 31 May 2022.
29. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet* **2020**; 395:470–473.
30. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* **2017**; 22:30494.

31. World Health Organization. COVID-19: variants. **2022**; Available at: <https://www.who.int/westernpacific/emergencies/covid-19/information/covid-19-variants>. Accessed 21 May 2022.
32. Kyriakidis NC, López-Cortés A, González EV, Grimaldos AB, Prado EO. SARS-CoV-2 vaccines strategies: a comprehensive review of phase 3 candidates. *NPJ Vaccines* **2021**; 6:28.
33. US Food & Drug Administration. Coronavirus Treatment Acceleration Program (CTAP). **2022**; Available at: <https://www.fda.gov/drugs/coronavirus-covid-19-drugs/coronavirus-treatment-acceleration-program-ctap>. Accessed 31 May 2022.
34. 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* **2022**; :e0020522.
35. 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.
36. 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. *Clin Infect Dis* **2021**; :ciab283.
37. 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.
38. Liu J, Laurie MT, Rubio L, et al. SARS-CoV-2 transmission dynamics and immune responses in a household of vaccinated persons. *Clin Infect Dis* **2022**; :ciac029.

2 GENOME-WIDE KNOCKOUT SCREEN IDENTIFIES HUMAN SIALOMUCIN CD164 AS AN ESSENTIAL ENTRY FACTOR FOR LYMPHOCYTIC CHORIOMENINGITIS VIRUS

2.1 ABSTRACT

Lymphocytic choriomeningitis virus (LCMV) is a well-studied mammarenavirus that can be fatal in congenital infections. However, our understanding of LCMV and its interactions with human host factors remain incomplete. Here, host determinants affecting LCMV infection were investigated through a genome-wide CRISPR knockout screen in A549 cells, a human lung adenocarcinoma line. We identified and validated a variety of novel host factors that play a functional role in LCMV infection. Among these, knockout of the sialomucin CD164, a heavily glycosylated transmembrane protein, was found to ablate infection with multiple LCMV strains but not other hemorrhagic mammarenaviruses, in several cell types. Further characterization revealed a dependency of LCMV entry on the cysteine-rich domain of CD164, including a N-linked glycosylation site at residue 104 in that region. Given the documented role of LCMV with respect to transplacental human infections, CD164 expression was investigated in human placental tissue and placental cell lines. CD164 was found to be highly expressed in the cytotrophoblast cells, an initial contact site for pathogens within the placenta, and LCMV infection in placental cells was effectively blocked using a monoclonal antibody specific to the cysteine-rich domain of CD164. Together, this study identifies novel factors associated with LCMV infection of human tissues, and highlights the importance of CD164, a sialomucin that has previously not been associated with viral infection.

2.2 IMPORTANCE

Lymphocytic choriomeningitis virus (LCMV) is a human pathogenic mammarenavirus that can be fatal in congenital infections. Although frequently used in the study of persistent infections in the field of immunology, aspects of this virus's life cycle remain incomplete. For example, while viral entry has been shown to depend on a cell adhesion molecule, DAG1, genetic knockout of this gene allows for residual viral infection, implying that additional receptors are able to mediate cell entry. The significance of our study is the identification of host factors important for successful infection, including the sialomucin CD164, which had not been previously associated with viral infection. We demonstrated that CD164 is essential for LCMV entry into human cells and can serve as a possible therapeutic target for treatment of congenital infection.

2.3 INTRODUCTION

The Arenaviridae family is classified into four genera: *Antennavirus* which were discovered in actinopterygian fish; *Reptarenavirus* and *Hartmanivirus* which infect boid snakes; and *Mammarenavirus* whose hosts are predominantly rodents [1–4]. Mammarenavirus can be further divided into two major virus subgroups based on antigenic properties: Old World (OW), which are mainly indigenous to Africa, and New World (NW), which are indigenous to the Americas [5]. Several viruses from this genus can also infect humans, leading to severe or fatal disease. One such pathogenic mammarenavirus is lymphocytic choriomeningitis virus (LCMV). Considered to be the prototypic arenavirus, LCMV is an OW virus found on all populated continents due to the ubiquitous distribution of its natural host, the house mouse (*Mus musculus*) [6]. The prevalence, however, among humans as measured through serological presence of LCMV antibodies widely varies (from 4% to 13%), making it challenging to estimate disease burden and infection risk [7]. In addition to contact with infected rodents, humans can

also become infected with LCMV through solid organ transplant or by vertical transmission. In the former case, LCMV infection in immunosuppressed organ recipients is frequently fatal and the only available therapeutic is off-brand use of nucleoside analog ribavirin [8]. As for the latter, transplacental infection leading to congenital LCMV are typically abortive or result in severe and often fatal fetopathy [9, 10].

Like all other mammarenaviruses, LCMV is a pleiomorphic enveloped virus with a bi-segmented, ambi-sense, negative-stranded RNA genome encoding four genes [11]. The L segment (7.2 kb) encodes the viral RNA-dependent RNA polymerase (L), and a small RING finger protein (Z) that is functionally equivalent to the matrix protein found in many enveloped RNA viruses. The S segment (3.4 kb) encodes the viral nucleoprotein (NP) and the glycoprotein complex (GPC). The GPC is synthesized in the infected cell as a precursor polypeptide before being proteolytically processed into a stable signal peptide (SSP), and two noncovalently linked subunits GP1 and GP2 by the protease SKI-1/S1P [12]. GP1 subunit associates with a cellular receptor while GP2 is a transmembrane protein that mediates the pH dependent fusion of viral and cellular membranes in the late-stage endosomes [13–15]. All three subunits remain associated in a tripartite complex while expressed on the viral surface to facilitate viral attachment and entry [12, 16].

Dystroglycan (DAG1), a widely expressed cell adhesion molecule, is recognized as the main attachment factor for viral entry by LCMV, Lassa Virus (LASV) and several other NW mammarenaviruses [17, 18]. DAG1 is expressed as a precursor polypeptide that is post-translationally cleaved into two noncovalently associated subunits, the peripheral membrane alpha subunit (α DG) and the transmembrane beta subunit (β DG) [19]. Additionally, α DG undergoes complex O-glycosylation mediated by the glycotransferase like-acetylglucosaminyl-transferase (LARGE). Appropriate LARGE-dependent glycosylation is critical for interaction between α DG and mammarenavirus GP [20, 21].

LCMV cellular tropism, however, does not always correlate with the presence of fully glycosylated α DG and certain strains of LCMV are still found to efficiently bind and infect host cells in the complete absence of DAG1 [22]. Previous studies have shown that single amino acid substitutions such as S153F, Y155H, and L260F in the GP1 domain can alter the binding affinity to α DG and shift GP binding preference to alternative receptors [23]. This allowed for further classification of LCMV strains into high- and low- α DG LCMV variants. Several secondary receptors have been proposed, including members of the Tyro3/Axl/Mer (TAM) family and heparan sulfate proteoglycans [24–26]. Interestingly, in each case, residual viral infection is still observed in when tested in genetic knockouts, implying the presence of additional receptors able to mediate cell entry.

The cell entry process reaches completion for mammarenaviruses when viral and cell membrane fusion allows the viral RNP to be deposited into the cytoplasm. For OW mammarenaviruses Lassa virus and Lujo virus, this step requires GP2 to bind to late endosomal resident proteins LAMP1 and CD36, respectively, in a low pH environment [27, 28]. Whether LCMV also requires such a receptor switch in the late endosome is currently unknown.

Although LCMV is considered the prototypic mammarenavirus and is consistently used as a model to study the effect of viral persistence on host immunity, several aspects of its viral life cycle and cellular tropism remain incompletely understood. In this study, we explored the essential host requirements for LCMV infection by performing a genome-wide CRISPR Cas9 knockout (KO) screen using the GeCKOv2 guide library [29]. Our results identify new host factors associated with LCMV infection, while also corroborating previously implicated factors. Among these results, we identify CD164 as an essential entry factor and possible therapeutic target for LCMV infection.

2.4 RESULTS

2.4.1 CRISPR KO screens identify host factors for LCMV infection

LCMV is a virus with minimal cytopathic effect. To conduct a genome-wide pooled CRISPR KO screen to identify host factors important for LCMV infection, a recombinant tri-segmented LCMV reporter virus (rLCMV-mCherry) with one L segment and two S segments was constructed [30]. We genetically encoded mCherry in place of the nucleoprotein (NP) on one S segment and in place of the glycoprotein complex (GPC) on the other S segment (**Figure 2.S1A**). One-step growth curves demonstrated slower growth kinetics for rLCMV-mCherry compared to its parental strain, LCMV Armstrong 53b (Arm 53b), with final titers being comparable (**Figure 2.S1B**). 24 hours post infection (hpi), the percentage of cells expressing mCherry (94.2% mCherry+) was equivalent to the percent expressing nucleoprotein (99.6% N protein+) suggesting minimal deleterious effects of this tri-segmented genome arrangement (**Figure 2.S1C**).

As inhalation of aerosolized virus is a major transmission route, human adenocarcinoma lung epithelial cells (A549) were the chosen cell line for whole-genome CRISPR KO screening with the GeCKOv2 guide library. Following rLCMV-mCherry infection (Multiplicity of infection (MOI) 10) of the A549 CRISPR KO library cells, mCherry-negative cells were sorted 24 hours post infection (hpi) to select for single-guide RNAs (sgRNAs) targeting host factors necessary for successful LCMV infection (**Figure 2.1A**). The sgRNAs present in this virus-resistant and an unsorted control population were PCR amplified from the extracted genomic DNA and subsequently identified via next-generation sequencing. Using the MAGeCK algorithm, genes were ranked using robust rank aggregation to produce a significance score called the MAGeCK enrichment score [31]. As expected, multiple sgRNAs targeting the same gene were among the top scoring guides, including those targeting previously described mammarenavirus host factors

(**Figure 2.1B**, **Figure 2.S2A**, **Table 2.S1**). These include sialic acid metabolism genes (*ST3GAL4*, *SLC35A1*) and glycosylation related genes (Conserved Oligomeric Golgi (COG) complex members, *TMEM165*) which have been shown to be LASV host factors [32]. Multiple heparan sulfate biosynthetic genes (*EXTL3*, *NDST1*, *PTAR1*, *SLC35B2*) described to be relevant for Lujo virus and DAG1-independent LCMV infections were also enriched [25, 28]. The LCMV attachment factor *DAG1* was detected, albeit at a lower enrichment score. Additional host factors that were significantly enriched include those described for other viral infections, such as negative-stranded RNA virus vesicular stomatitis virus (VSV) (*ARFRP1*, *SYS1*, *YKT6*) and the human immunodeficiency virus (HIV) (*SRP14*, *DYRK1A*, *IL2RA*) [33–37].

Gene Ontology (GO) overrepresentation analysis of the top 300 hits from the screen using PANTHER [38] indicated an enrichment of genes associated with the signal recognition particle (*SRP14*, *SRP68*, *SRP19*, *SRP72*) and proton transmembrane transporter activity (*ATP6V1E1*, *ATP6V0D1*, *ATP6V1B2*, *ATP6V0C*, *ATP6V1A*, *ATP6V1G1*, *ATP6V0B*, *ATP12A*, *ATP6V1H*, *ATP6V1C1*, *CLCN4*, *ATP6V1F*, *ATP5S*) (**Table 2.S2**). These same hits are also overrepresented in GO cellular components signal recognition particle and vacuolar proton-transporting vacuolar type ATPase (v-ATPase) complex, respectively.

Nearly every subunit of the v-ATPase (*ATP6V1B2*, *ATP6V0C*, *ATP6V0B*, *ATP6V1G1*, *ATP6AP1*, *ATP6V0D1*, *ATP6V1C1*, *ATP6V1E1*, *ATP6V1H*, *ATP6AP2*) was enriched in our screen. V-ATPase is a proton pump responsible for acidification of intracellular systems, a process necessary for the required pH-dependent fusion event between LCMV viral and cellular membranes in the acidic environment of the late-stage endosome [15, 39]. To validate this screening result, known v-ATPase inhibitors Bafilomycin A₁ (**Figure 2.S2B**), Bafilomycin B₁ (**Figure 2.S2C**), and Concanamycin A (**Figure 2.S2D**) were tested for efficacy in LCMV infection inhibition [40]. As expected, all three drugs exhibited dose-dependent protection

against LCMV infection in A549 cells with nanomolar efficacy, consistent with the critical role v-ATPase plays in LCMV infection.

To explore other candidate genes of interest identified in this screen, monoclonal A549 knockout (KO) cell lines containing frameshift mutations were generated for several top-scoring genes (**Figure 2.S2E**). These cell lines were also tested for normal cell growth (**Figure 2.S2F**). Among these candidates were the transmembrane proteins encoded by *ACKR4*, *CD164*, *EMC1*, *IL2RA*; the trans-Golgi/endosome membrane trafficking complex *ARFRP1* and *SYS1*; the vesicular transport associated genes *YKT6* and *RAB10*; the ZAP anti-viral protein co-factor *KHNYN*; and the signal recognition particle gene *SRP14*. In all cases, homozygous and heterozygous knockouts in A549 cells yielded significant decreases in LCMV infection, ranging from severely impaired relative to wildtype: 1.3% infected (*-/-CD164*), to moderately impaired: 77% infected (*-/+SYS1*). (**Figure 2.1C**)

Since knockout of CD164 demonstrated near ablation of infection, we chose to follow-up on this protein to explore its role in the viral life cycle. CD164 is a heavily glycosylated transmembrane sialomucin cell adhesion protein expressed in a wide range of tissues [41, 42]. This gene was originally characterized as a marker for CD34+ hematopoietic progenitor cells where it may be involved in a variety of processes, including cellular adhesion, autophagy, tumorigenesis, and metastasis [43, 44]. To date, *CD164* has not been associated with any known viral entry mechanisms.

To further investigate the role of this gene in LCMV infection, monoclonal *CD164* KO (Δ *CD164*) cell lines were generated in two additional cell types: 293T (human embryonic kidney cells) and 3T6-Swiss albino (mouse embryonic fibroblast cells). In both human lines, A549 and 293T, deletion of CD164 reduced infection by 99% and 95% respectively, while the effect in the mouse cell line 3T6 was moderate (41% reduction) (**Figure 2.1D**). Infectivity of each KO cell line was nearly fully restored by complementation with ectopically expressed human *CD164*

(*hCD164*) gene driven by the CMV promoter. Complementation with the mouse *Cd164* (*mCD164*) gene, which is 62.32% identical on a protein level, partially restored infectivity in all three cell lines. We confirmed protein expression levels in knockout and complemented cell lines by Western blot (**Figure 2.S2G-I**).

In addition to viral infections, one-step growth curves were also conducted to determine the abundance of infectious viral progeny produced by $\Delta CD164$ cells (**Figure 2.S2J**). As expected, few productive virions were produced from $\Delta CD164$ A549 cells when compared to the WT A549 cells. Together, our data suggests that *CD164* is essential for LCMV infection in human cells.

2.4.2 Pseudotyped viral infection shows that CD164 is a LCMV-specific mammarenavirus human entry factor

Previous work has demonstrated *DAG1* to be an entry-related attachment receptor in mice [17]. Our screen also identified *DAG1* as an important LCMV entry factor in addition to implicating *CD164* as a determinant of human cell entry for LCMV. To further explore the dependency of mammarenaviruses on *CD164* or *DAG1* for viral entry, we generated and validated additional monoclonal KO cell lines, $\Delta DAG1$ and $\Delta CD164/\Delta DAG1$ double KO, in both A549 (**Figure 2.S3A**) and 293T (**Figure 2.S3B**) cell backgrounds. To specifically test the entry stage of the viral life cycle, recombinant green fluorescent protein (GFP) expressing vesicular stomatitis virus (rVSV- ΔG (GFP)) pseudotyped with a panel of mammarenavirus GP were utilized [45]. The advantage of this method is two-fold: targeted examination of GP receptor tropism in the absence of other factors that may influence native viral infection and the ability to study BSL-4 pathogenic mammarenaviruses in standard BSL-2 laboratory conditions [46].

The GPs from several LCMV strains representing a range of DAG1 affinities were combined with rVSV- Δ G(GFP) to generate pseudotyped virus (**Figure 2.2A-E**). Arm 53b (used for the CRISPR screen) (**Figure 2.2A**) and WE2.2 (**Figure 2.2B**) represent low DAG1 affinity strains, while Armstrong Clone 13 (Arm C113) (**Figure 2.2C**), WE54 (**Figure 2.2D**), and WE (**Figure 2.2E**) were chosen to represent high-DAG1 affinity strains [22, 23, 25]. Deletion of *CD164* reduced infection by all four pseudotyped viruses by 78%-99% in both human cell lines, indicating a strong *CD164* dependency in all cases. In contrast, knockout of *DAG1* in both A549 and 293T cells led to only moderate decreases in pseudotyped virus infection (23%-38% reduction for A549; up to 63% reduction for 293T) across all LCMV strains. In one case (293T Δ *DAG1* infected with rVSV- Δ G(GFP)+WE54-GP), deletion of *DAG1* had virtually no measurable impact on pseudotyped virus infection. Consistent with that, Δ *DAG1*/ Δ *CD164* double KO cells yielded decreased pseudovirus infection similar or below those observed in *CD164* KO cells. Together, these results suggest that *CD164* is the major determinant for LCMV entry in both A549 and 293T cells, whereas *DAG1* plays only an accessory role in these human cell types.

To extend these findings beyond LCMV, a selection of hemorrhagic mammarenaviruses GPCs were used to generate pseudovirus for infection in A549 and 293T cells. As previously described, *DAG1* is an important entry factor for the OW mammarenavirus Lassa Virus (LASV), but not for the NW mammarenaviruses, Guanarito virus (GTOV) and Machupo virus (MACV) [17, 18]. Consistent with these findings, deletion of *DAG1* abrogated LASV entry (**Figure 2.2F**) but had minimal effects on GTOV (**Figure 2.2G**) and MACV entry (**Figure 2.2H**). By contrast, deletion of *CD164* had no effect on infection for any of the three tested pathogenic mammarenaviruses, suggesting that *CD164* is a critical human entry factor for LCMV, but not other mammarenaviruses. Using GP-pseudotyped virus, we have determined that *CD164* plays a major functional role for LCMV entry in human cells and no other tested hemorrhagic

mammarenaviruses, while *DAG1* is an important entry factor for LASV and, to a lesser extent, certain strains of LCMV.

2.4.3 N-linked glycosylation within the cysteine-rich domain is critical for LCMV infection

CD164 is a 197 amino acid type 1 integral transmembrane protein featuring a 14 amino acid intracellular tail and a 139 amino acid extracellular region that is expressed as a homodimer nearly ubiquitously throughout human tissues [47]. The extracellular portion of *CD164* is comprised of two mucin domains flanking a cysteine-rich domain. The protein also features one predicted attachment site for O-linked glycans and 9 predicted N-linked glycosylation sites throughout the mucin and cysteine-rich domains.

To further dissect the role of *CD164* with respect to LCMV entry, a series of *CD164* domain deletion mutants were constructed and introduced into A549 $\Delta CD164$ and 293T $\Delta CD164$ cells (**Figure 2.3A**). Deletion of the first mucin domain ($\Delta CD164 + hCD164(\Delta E1)$) did not affect infection, suggesting this domain is not necessary for LCMV entry. Extending the deletion into the cysteine-rich domain ($\Delta CD164 + hCD164(\Delta E1-2)$) however, ablated infection (mean 98% reduction for A549; 86% reduction for 293T), thereby phenocopying the $\Delta CD164$ cells. We confirmed expression of all domain deletion constructs by Western blot (**Figure 2.S4A and 2.S4B**).

The cysteine-rich region of *CD164* contains four putative N-linked glycosylation sites (**Figure 2.3B**). To test the importance of these sites individually, alanine substitutions were introduced in place of each relevant asparagine, and expression of these mutant construct was confirmed by Western blot (**Figure 2.S4C and 2.S4D**). Mutation of N-linked glycosylation sites at positions 72, 77, and 94 did not reduce infection by rLCMV-mCherry, however, substitution of N104 completely abolished infection. This asparagine residue, which is conserved between

human and mouse CD164 (**Figure 2.3C**), sits in a loop region between a beta-sheet and an alpha-helix as predicted by AlphaFold (**Figure 2.S4E**) [48]. The ablation of infection due to mutagenesis of the N-linked glycosylation site suggests that the cysteine-rich domain, including a critical asparagine amino acid, is required for *CD164*-mediated infection by LCMV.

The deletion mapping of *CD164* indicated the importance of the cysteine-rich domain. To further explore this domain, we tested whether an anti-CD164 monoclonal antibody (mAb) could competitively inhibit LCMV infection. The anti-CD164 mAb N6B6, which was demonstrated to bind a conformationally dependent backbone epitope encompassing the cysteine-rich domain between the two mucin domains [42, 49], blocked infection by rLCMV-mCherry in a dose-dependent manner (**Figure 2.3D**). These results are consistent with the deletion mapping and alanine mutagenesis data, highlighting the importance of the central cysteine-rich domain for LCMV infection.

2.4.4 CD164 is highly expressed in human placenta and mediates LCMV infection in placental cells

Although LCMV infection as a child or an adult are typically inconsequential, infection during pregnancy can lead to transplacental human fetal infections with severe clinical consequences [10]. Like many other congenital pathogens, LCMV has tropism for fetal neural and retinal tissue, leading to developmental issues such as microencephaly, macrocephaly, chorioretinitis, periventricular calcification, and hydrocephalus [50, 51]. Retrospective studies on serologically confirmed cases show that children with congenital LCMV infection have a 35% mortality rate by 2 years of age and survivors experience long-term neurological, motor, and visual impairments [9, 52].

Human fetal vulnerability to LCMV led us to hypothesize that CD164 may play a role in transplacental infection. To explore tissue specific expression of CD164 during pregnancy, healthy second trimester placentas were co-stained with CD164 mAb anti-CD164 N6B6. CD164 was highly expressed in the outer layer of floating chorionic villi and absent in the underlying mesenchyme; co-localization with cytokeratin-7 confirmed that CD164 was expressed in cytotrophoblasts (**Figure 2.4A**) [53]. In contrast, CD164 was not detected in the decidua (maternal side), pointing to a fetal-specific localization at this interface (**Figure 2.S5A**). Cytotrophoblasts bathe in maternal blood and are an initial contact site for pathogens [54], suggesting that CD164 is present in tissue structures and locations amenable for transplacental infection of the developing fetus.

2.5 DISCUSSION

In this study, we performed a genome-wide CRISPR KO screen to identify host factors important for LCMV. By using a reporter virus, our screen is well-suited to specifically identify host factors involved in viral entry and viral protein production. Furthermore, we chose to conduct our screen using a high MOI to decrease the probability of enriching sgRNAs that may be present in our sorted uninfected population due to chance. Our results ultimately highlight a subset of genes that appear to be shared generally among mammarenaviruses. These genes span pathways and functions such as sialic acid metabolism, heparan sulfate biosynthesis, glycosylation and Golgi trafficking, and late-stage endosome acidification [25, 28, 32]. Most notably, we identified 10/24 of the v-ATPase subunits (5 in each of the V_1 and V_0 domains) as well as four signal recognition particle subunits. The previously described LCMV entry receptor, *DAG1*, was moderately enriched, consistent with the use of the low DAG1 affinity Arm 53b LCMV strain in this screen [22, 23].

This screen also revealed genes, most notably CD164, that have not previously linked to LCMV infection, perhaps facilitated by using a human epithelial lung cell line. We found that in the absence of CD164 in human cells LCMV infection is nearly ablated. In contrast, mouse $\Delta CD164$ cells yielded only moderate reduction in LCMV infection. Consistent with this, when complemented with ectopically expressed human or mouse CD164, human CD164 restored infectivity, while the mouse homologue of CD164, which is 62% identical on the protein level, only partially restored infectivity. CD164 localizes to the cell surface and late-stage endosomes, consistent with the LCMV entry route for successful infection [47]. Like DAG1, it is a ubiquitously expressed cell adhesion molecule present in nearly all tested human tissue. Unlike DAG1, to which LCMV strains show a range of affinity, all five LCMV strains tested here required CD164 for infection in human cells. Deletion of DAG1 partially reduced infectivity by LCMV, particularly in those previously described as having high DAG1 affinity (CI13, WE54 and WE). These data strongly support that human lung cells require CD164, and not necessarily DAG1, for viral infection by LCMV, whilst mouse cells appear to rely on CD164 only partially.

Further characterization of *CD164* by deletion mapping and alanine mutagenesis suggests that the cysteine-rich domain, particularly a single critical N-linked glycosylation site, is required for CD164-mediated infection. The importance of the cysteine-rich domain was reinforced by blocking using the anti-CD164 mAb N6B6, whose presence can inhibit LCMV infection in a dose-dependent manner. These data together suggest that binding by N6B6 to CD164 renders the critical interaction region inaccessible, and thus preventing LCMV infection.

While LCMV infection is generally mild among adults and children, clinical outcomes following congenital infections tend to be severe. LCMV transplacental infections are typically fatal and survivors experience long-term neurological, motor, and visual impairments [9, 10, 52]. While off-brand use of ribavirin occurs in cases of LCMV infection following solid organ transplants, no current treatment procedure exists for congenital LCMV [8]. We demonstrated

that CD164 localizes to the placental villous cytotrophoblasts, the maternal-fetal interface that is the initial contact site with congenital pathogens. The function of this sialomucin in placentation is unknown and, to our knowledge, this is the first description of CD164 expression in human placenta. Placental sialoglycoproteins disarm maternal immune recognition of fetal antigens [55], thus making CD164 a particularly elegant target for viral entry. LCMV may evade immunity at this critical interface by co-opting natural mechanisms of maternal immune tolerance of her fetus. Once in the fetus, LCMV exhibits brain tropism, specifically targeting neuroblasts. The developing brain is most susceptible as more severe pathology is observed when infection occurs early in gestation, possibly due to incomplete blood-brain barrier formation [51]. LCMV also induces fetal inflammation, which is highly damaging and can lead to inflammation-induced cerebral lesions in children with congenital infection [51, 56]. Thus, either directly or indirectly, the ability of LCMV to enter the protected fetal compartment likely via engagement of CD164 contributes to the devastating neurological pathology associated with the virus. Our identification of a key placental entry factor for LCMV and demonstration that an antibody can inhibit this interaction, opens the possibility of targeting this pathway for therapeutic intervention.

While this study was under review, a report by Bakkers et al. [57] was published detailing the role of CD164 in relation with LCMV-GP during the viral entry process. Using a LCMV pseudotyped virus and an infection period of 3 weeks, the authors identified seven host factors important for LCMV entry, four of which (CD164, ST3GAL4, SLC35A1, MAN1A2) were contained within the set of 37 factors identified in our live virus screen (**Figure 2.S6**). CD164 was also demonstrated to be required for low-pH triggered membrane fusion. Similar to LASV, which binds to DAG1 at the cell surface and subsequently uses a receptor switching strategy to transition to LAMP1 binding in the acidic environment of the late-stage endosome [27], LCMV is believed to use a similar mechanism by changing receptors to CD164 for viral membrane fusion.

The data presented in this study are consistent and complementary to that presented by Bakkers et al. In addition, these data extend identification of host factors important for LCMV infection to 33 additional genes, including members of complexes such as the v-ATPase. Furthermore, we have demonstrated that a monoclonal antibody, N6B6, is able to bind to CD164 and inhibit LCMV infection in a dose-dependent manner. In addition to lung epithelial cells, this antibody is also shown to be effective in preventing infection in placental cells, thereby presenting a possible therapeutic for preventing congenital LCMV infection.

The demonstration of CD164 as an essential determinant for LCMV entry into human cells and tissue fills an important gap in our understanding of this virus. Whether the reliance on CD164 is unique to LCMV, or whether this entry factor is utilized by additional viruses remains unknown. Given the apparent unique dependency of LCMV on CD164, and the practical implications in its involvement in transplacental infection, further exploration of the mechanistic details by which congenital LCMV can be prevented through blocking CD164 is warranted.

2.6 MATERIALS AND METHODS

2.6.1 Cell lines

A549 (ATCC), 293T (ATCC), 3T6 (ATCC), BHK-21 (ATCC), and Vero cells (ATCC) were cultured in DMEM (Gibco) supplemented with 10% fetal bovine serum (FBS, Gibco), penicillin-streptomycin-glutamine (Gibco), and HEPES (Gibco) at 37°C and 5% CO₂. JEG-3 (ATCC) were cultured in EMEM (ATCC) supplemented with 10% FBS (Gibco), penicillin-streptomycin-glutamine (Gibco), non-essential amino acids (Gibco), and sodium pyruvate (Gibco) at 37°C and 5% CO₂. All cell lines tested negative for mycoplasma contamination (Lonza).

2.6.2 Genome-wide CRISPR screen

A549 cells were stably lentivirally transduced with Cas9-Blast (Addgene #52962, gift from Feng Zhang) and subsequently selected using blasticidin. Next, a total of 300 million A549-Cas9 cells were then transduced with the lentiviral human GeCKO v2 library (Addgene #1000000049, gift from Feng Zhang) [29] at a MOI 0.5 and selected using puromycin for 6 days. To conduct the host factor screen, 120 million (60 million each of sub-library A and B) A549-Cas9-Blast GeCKO library cells were infected with rLCMV at MOI 10. At 24 hpi, cells that remained mCherry-negative were collected using a Sony SH800 cell sorter. Simultaneously, 120 million cells of uninfected A549-Cas9-Blast GeCKO library cells were collected to assess sgRNA representation as a reference.

Genomic DNA (gDNA) was extracted using the NucleoSpin Blood kit (Macherey-Nagel). The sgRNA expression cassettes were amplified from gDNA in a two-step nested PCR using Q5 High-Fidelity 2X Master Mix (NEB). For PCR-I, 48 reactions (for control samples) and 12-24 reactions (for mCherry-negative sorted FACS samples) containing 1 μ g were amplified for 16 cycles. Reactions were pooled, mixed and size-selected using SPRIselect (Beckman Coulter). During PCR-II, 10 reactions containing 5 μ L of PCR-I product were amplified for 10 cycles using indexed Illumina primers. PCR products were cleaned using AmpureXP beads (Beckman Coulter) and sequenced on an Illumina NextSeq 500 using a custom sequencing primer. Primer sequences can be found in **Table 2.S4**. Raw sequencing reads can be found at NIH BioProject Accession Number: PRJNA806912.

Demultiplexed FASTQ files were aligned to a reference table containing sgRNA sequences and the abundance of each sgRNA was determined for the starting and sorted cell population. Guide count tables were further processed using MAGeCK to determine positive enrichment scores for each gene [31]. Gene ontology enrichment was determined with

statistical overrepresentation test on PANTHER [38] using genes from the 300 highest MAGeCK scores.

2.6.3 Generation of monoclonal KO cell lines

sgRNA sequences against gene targets were designed using CRISPick [59] and the corresponding DNA oligos (IDT) were annealed and ligated into pX458 (Addgene #48138, gift from Feng Zhang) [60]. Cells were transfected with pX458 constructs using *TransIT-X2* (Mirus Bio) and GFP positive cells were sorted into 96-well plates using a FACS Aria II (BD) two days later. Clonal populations were genotyped by Sanger sequencing the PCR amplified sgRNA-targeted sites in the gDNA extracted using DNA QuickExtract (Lucigen). Resulting sequences were compared to references and clones containing a frameshift indel were selected. To determine cell growth of A549 WT and KO cell lines, CellTiter-Glo (Promega) was mixed 1:1 with cells seeded in 96-well plates for three consecutive days and the luminescence signal was quantified using the GloMax-Multi microplate reader (Promega). A list of all used sgRNA sequences and genotyping primers can be found in **Table 2.S4**.

2.6.4 Plasmids, cloning, and lentivirus production

Human CD164 (Origene, #RC202234) and mouse Cd164 (Origene, #MR201951) cDNAs were cloned into EcoRV-cut plenti-CMV-Puro-DEST (Addgene #17452, gift from Eric Campeau & Paul Kaufman) [61] using NEBuilder HiFi DNA Assembly Master Mix (NEB). Primers used to assemble expression plasmids for domain deletion mapping and alanine scanning mutagenesis of CD164 can be found in **Table 2.S4**.

Lentivirus was produced in HEK293T by co-transfection of cDNA containing lentiviral plasmid together with helper plasmids pMD2.G (Addgene #12259, gift from Didier Trono) and

pCMV-dR8.91 (Life Science Market) using *TransIT-Lenti* (Mirus Bio). Supernatant were collected 48h post-transfection, filtered, and added to recipient cells in the presence of Polybrene (EMD Millipore). Transduced cells were subsequently selected using Puromycin (Thermo Fisher Scientific) during days 3-5.

2.6.5 Compound inhibition and antibody neutralization

Bafilomycin A₁, Bafilomycin B₁ (Cayman Chemical Company), and Concanamycin A (Santa Cruz Biotechnology) were resuspended in DMSO and stored at -20C until use. Cells were incubated with compounds for 1h at 37C prior to infection assay.

Antibody neutralization assays were conducted by pre-incubating cells with anti-CD164 clone N6B6 (BD Pharmingen) or mouse IgG isotype control (BD Pharmingen) for 1h at 37C prior to infection assay. Antibody details can be found in **Table 2.S3**.

2.6.6 Generation of Arenavirus pseudotyped vesicular stomatitis virus

Glycoprotein from LASV (Genbank: AAA46286.1), GTOV (Genbank: AAN05423.1), MACV (Genbank: AIG51558.1), and LCMV strain WE-HPI (Addgene #15793, gift from Miguel Sena-Esteves) [62] were cloned into a pCAGGS vector backbone using NEBuilder HiFi DNA Assembly Master Mix (NEB). To generate an LCMV strain CI13-GP (Genbank: DQ361065.2) expression plasmid, mutations N176D and F260L were introduced to pCAGGS-LCMV-Arm4-GP using site-directed mutagenesis. To generate LCMV strain WE54-GP (Genbank: AJ297484.1), mutations V94A, S133T, Y155H, and T211A were introduced into LCMV strain WE-HPI-GP. To generate LCMV strain WE2.2-GP (Genbank: AJ318512.1), mutation S153F was introduced into LCMV strain WE54-GP [22, 23]. A list of primers used for cloning and site-directed mutagenesis can be found in **Table 2.S4**.

To rescue the various VSV- Δ G-Arenavirus-GP pseudotype virus, 293T cells were transfected with arenavirus glycoprotein expression plasmids using *TransIT-LT1* (Mirus Bio). Cells were transduced the following day with VSV- Δ G-GFP (Kerafast) [45] at MOI 3 and incubated in media containing anti-VSV-G antibodies (Kerafest) for 24h. Clarified supernatant containing pseudovirus were collected and stored at -80C. Stock titers were measured using flow cytometry on a FACSCelesta (BD). All experiments with pseudotyped VSV were performed in a BSL 2 laboratory.

2.6.7 Flow cytometry analysis of viral infection assays

Cells plated in 96-well plates were infected with rLCMV-mCherry or LCMV at MOI 1 for an adsorption period of 1h at 37C and subsequently cultured for 24h. To analyze percent infected, cells were trypsinized and fixed in suspension with 4% PFA for 30min. For infection with rLCMV-mCherry, analysis was done by flow cytometry on FACSCelesta (BD) where approximately 5,000 cells were recorded and gated based on SFC/SSC, FSC-H/FSC-A (singlets), and PE-CF594 (mCherry) using FlowJo 10. For infection with LCMV, cells were permeabilized and stained for LCMV-N protein (primary: 113, secondary: Alexa Fluor 488) prior to flow gating for FITC. Antibody details can be found in **Table 2.S3**.

For pseudotype infection assays, cells seeded in a 96-well plate were infected with various VSV-Arenavirus-GP pseudoviruses. At 24 hpi, cells were lifted using Tryple Select Enzyme (Gibco) and flowed on a FACSCelesta (BD), and for FITC (eGFP) signal as previously described.

2.6.8 Indirect Immunofluorescence of Human Placenta

Five micron cryosections of human placenta were post-fixed, permeabilized, blocked, and stained using standard protocols [53]. Briefly, primary CD164 mouse anti-human Clone N6B6 (BD Pharmingen), mouse isotype (BD Pharmingen), or Cytokeratin-7 rat anti-human (Sigma Millipore) were used to probe sections overnight at 4C. Immunofluorescence was detected using fluoroconjugated antibodies at 1:1000 dilution for 1h at room temperature. Nuclei were stained with NucBlue (Thermo Fisher) and slides were mounted in ProLong Gold Anti-fade reagent (Thermo Fisher) according to manufacturer's instructions. Slides were visualized using a Nikon Ti2 inverted fluorescence microscope with a Crest large field of view spinning disk confocal (CrestOptics). Antibody details can be found in **Table 2.S3**.

2.6.9 Western Blots

Cells were scraped and lysed in RIPA buffer on ice. All lysates were separated by SDS-PAGE on pre-cast 4-12% Bis-Tris gels (Thermo Fisher) in the NuPAGE electrophoresis system. Proteins were transferred onto nitrocellulose membrane using the Bio-Rad Mini-Protean Mini Trans-Blot transfer system. Membranes were blocked with Tris-buffered saline with 0.05% Tween-20 and 5% non-fat milk and incubated with primary antibody diluted in blocking buffer overnight at 4C on a shaker. Primary antibodies were detected by incubating membranes with 1:15,000 dilution of IRDye secondary antibodies (LI-COR) for 1h at room temperature and visualized using the Odyssey CLx (LI-COR). Antibody details can be found in **Table 2.S3**.

2.6.10 Quantification and Statistical Analysis

For viral infection, drug treatment, antibody neutralization, and cell growth experiments, biological replicates are defined as independent treatments and measurements from cells

harvested from multiple wells on different days. Replicates are displayed as mean \pm SEM and visualized using GraphPad Prism 9. Dose-response curves for drug treatments and antibody neutralizations were generated by applying a non-linear curve fit with least-squares regression and default parameters using GraphPad Prism 9. For all experiments, the statistical details can be found in the figure legends.

2.7 ACKNOWLEDGEMENTS

We would like to thank Dr. Stephanie Gaw for gifting placental specimens used in our microscopy studies and Dr. Michael Buchmeier for thoughtful discussions. This work was supported by the National Institutes of Health [F31AI150007 to S.S.]; the Chan Zuckerberg Biohub [to J.D., C.Y.W., and A.P.]; and the Chan Zuckerberg Initiative [to J.D.]. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

2.8 FIGURES

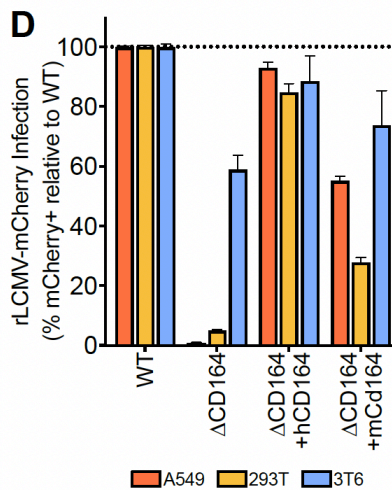
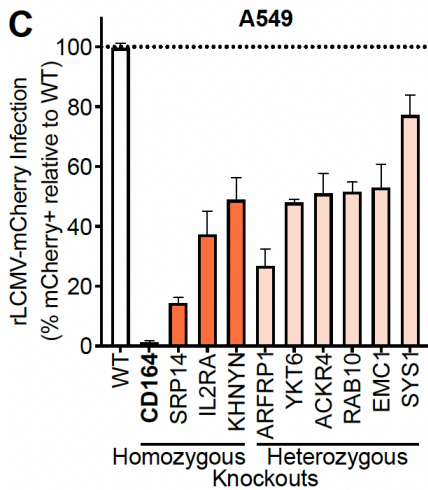
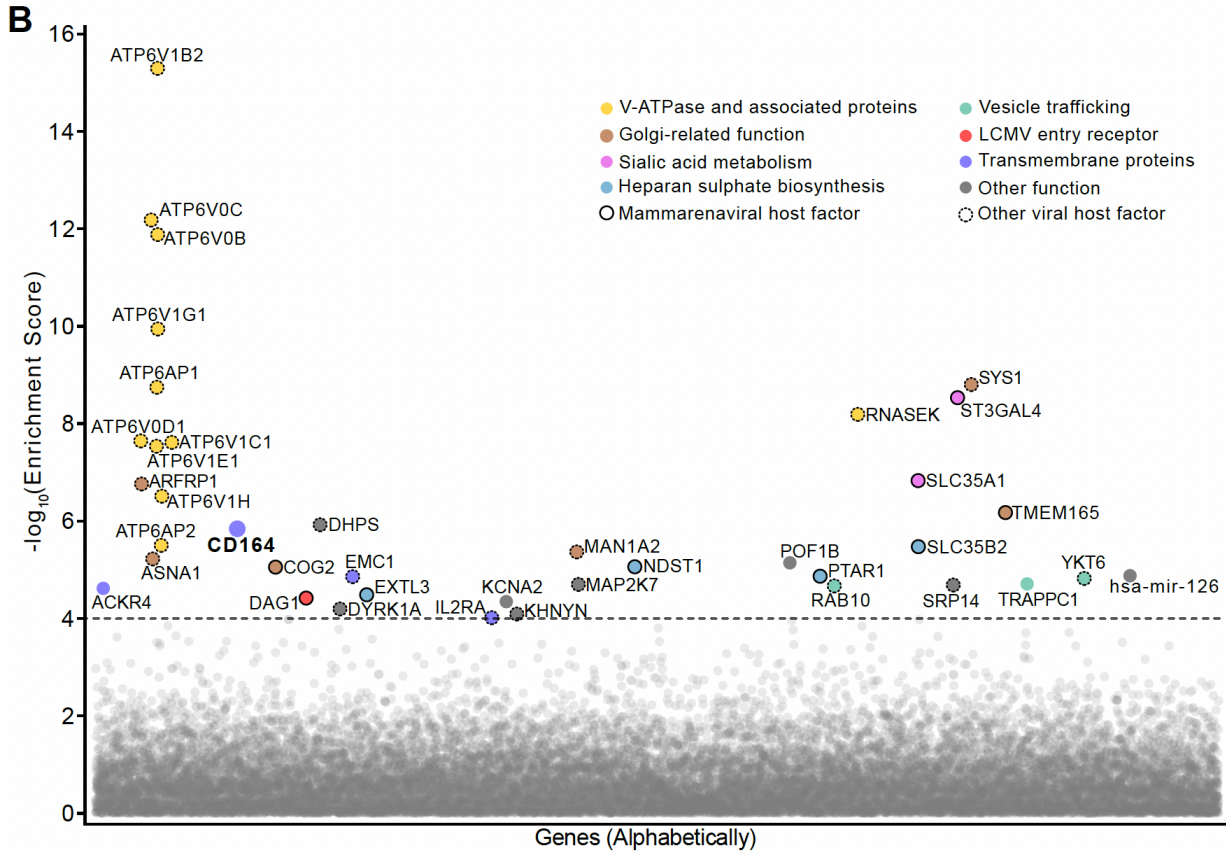
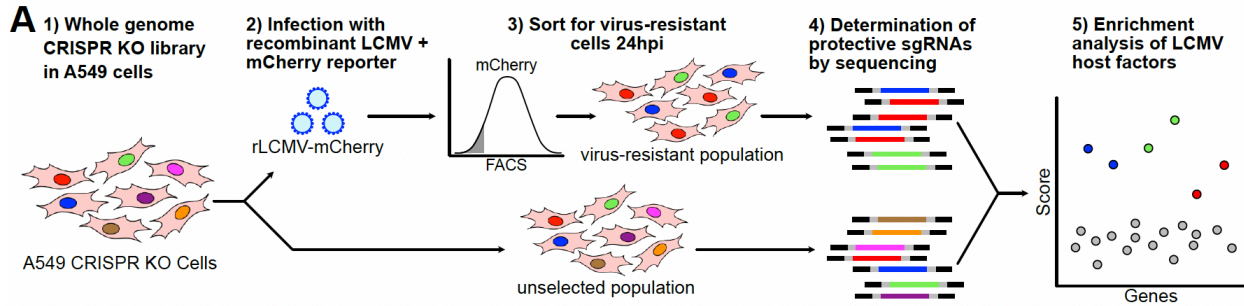


Figure 2.1 Genome-wide CRISPR loss-of-function screen in human cells identify host factors important for LCMV infection.

(A) Schematic of CRISPR-based KO screen done in A549 lung epithelial cells for the identification of LCMV host factors. (B) Gene enrichment for CRISPR screen of rLCMV-mCherry infection. Enrichment scores were determined by MaGECK analysis and genes were colored by biological function. Dotted line indicates $-\log_{10}(\text{Enrichment Score}) = 4$. All genes and their enrichment scores can be found in **Table S1**. (C) Percentage of infected cells as determined by flow cytometry following infection of A549 homozygous knockouts (CD164, SPR14, IL2RA, KHNYN) or heterozygous knockouts (ARFRP1, YKT6, ACKR4, RAB10, EMC1, SYS1) with rLCMV-mCherry. Wildtype cells were used as normalization controls. Cells were infected at MOI 1 and harvested at 24 hpi. Error bars indicate standard error of three independent experiments. (D) Quantification of viral infection in WT, $\Delta CD164$, $\Delta CD164$ complemented with human *CD164* ($\Delta CD164 + hCD164$), and $\Delta CD164$ complemented with mouse *Cd164* ($\Delta CD164 + mCd164$) in A549, 293T, and 3T6 cell type backgrounds. Cells were infected with rLCMV-mCherry at MOI 1 and harvested at 24 hpi. Error bars indicate standard error of three independent experiments.

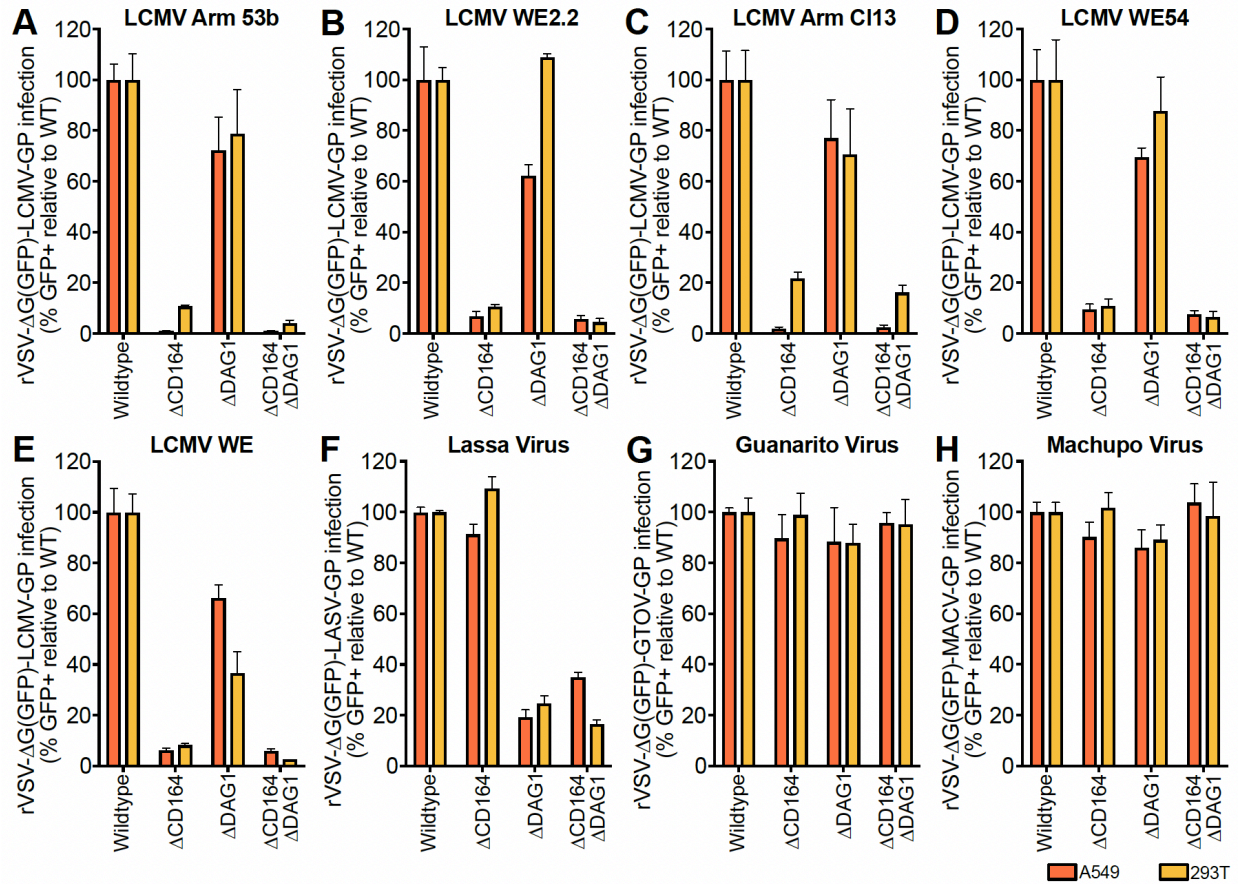


Figure 2.2 Infection of KO cell lines with a panel of mammarenavirus GP pseudotyped virus
 (A-E) Percent infection of $\Delta CD164$, $\Delta DAG1$, and $\Delta CD164/\Delta DAG1$ double-KO cells relative to WT in either A549 or 293T cell type backgrounds following inoculation with low DAG1 affinity LCMV strains (A) Armstrong 53b-GP or (B) WE2.2-GP, and high DAG1 affinity strains (C) Armstrong Clone 13-GP, (D) W54-GP, or (E) WE-GP pseudotyped virus as determined by flow cytometry for GFP positivity. Cells were infected at MOI 1 and measured 24 hpi. Error bars indicate standard error of three independent experiments. (F-H) Percent infection of $\Delta CD164$, $\Delta DAG1$, and $\Delta CD164/\Delta DAG1$ double-KO cells relative to WT in either A549 or 293T cell type backgrounds following inoculation with (D) LASV-GP, (E) GTOV-GP, or (F) MACV-GP pseudotyped virus as determined by flow cytometry for GFP positivity. Cells were infected at MOI 1 and measured 24 hpi. Error bars indicate standard error of three independent experiments.

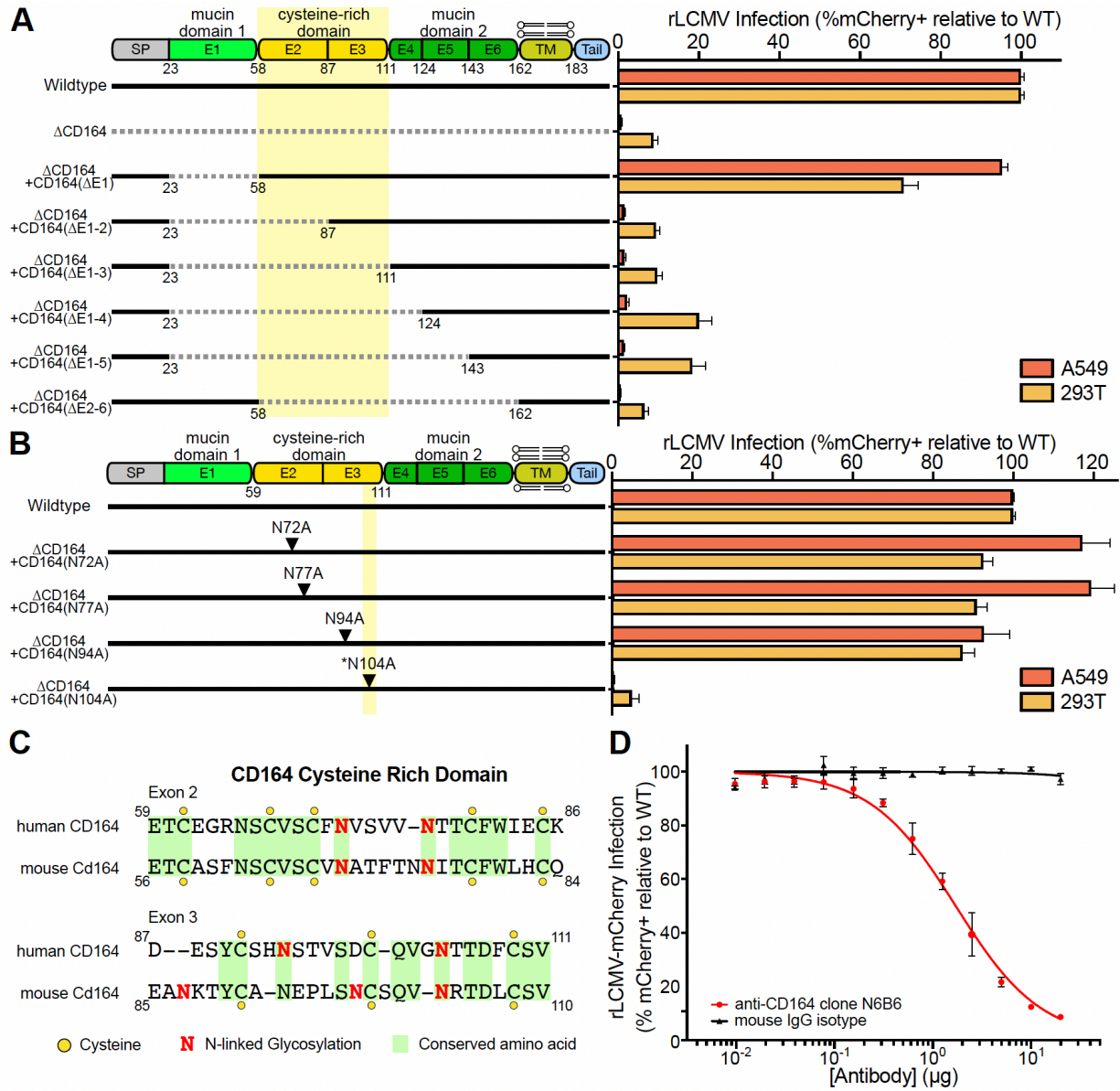


Figure 2.3 CD164 functional region determination through antibody binding, domain deletion, and alanine mutagenesis

(A) Schematic (left) of wildtype, Δ CD164, Δ CD164 + hCD164(Δ E1), Δ CD164 + hCD164(Δ E1-2), Δ CD164 + hCD164(Δ E1-3), Δ CD164 + hCD164(Δ E1-4), Δ CD164 + hCD164(Δ E1-5), and Δ CD164 + hCD164(Δ E2-6). Complemented A549 and 293T cells were challenged with rLCMV-mCherry (MOI 1) and infection was measured by flow cytometry at 24 hpi (right). Percent infection was normalized to wildtype. Error bars represent standard error of three independent experiments. (B) Schematic (left) of wildtype, Δ CD164 KO + hCD164(N72A), Δ CD164 + hCD164(N77A), Δ CD164 + hCD164(N94A), and Δ CD164 + hCD164(N104A). Complemented A549 and 293T cells were challenged with rLCMV-mCherry (MOI 1) and infection was measured by flow cytometry at 24 hpi (right). Percent infection was normalized to wildtype. Error bars represent standard error of three independent experiments. (C) Amino acid similarities of the cysteine-rich region in human CD164 and mouse Cd164 determined using the ClustalW program on SnapGene. Yellow circles indicate cysteine residues, red N symbolizes N-linked glycosylation sites, and identical amino acids are highlighted in green. (D) Blockade of LCMV infection with serial dilutions of anti-human CD164 monoclonal mouse antibody clone N6B6 or mouse IgG2a-k isotype control

in wild type A549 cells. Cells were infected at MOI 1 and infection measured at 24 hpi. Error bars indicate standard error of three independent experiments.

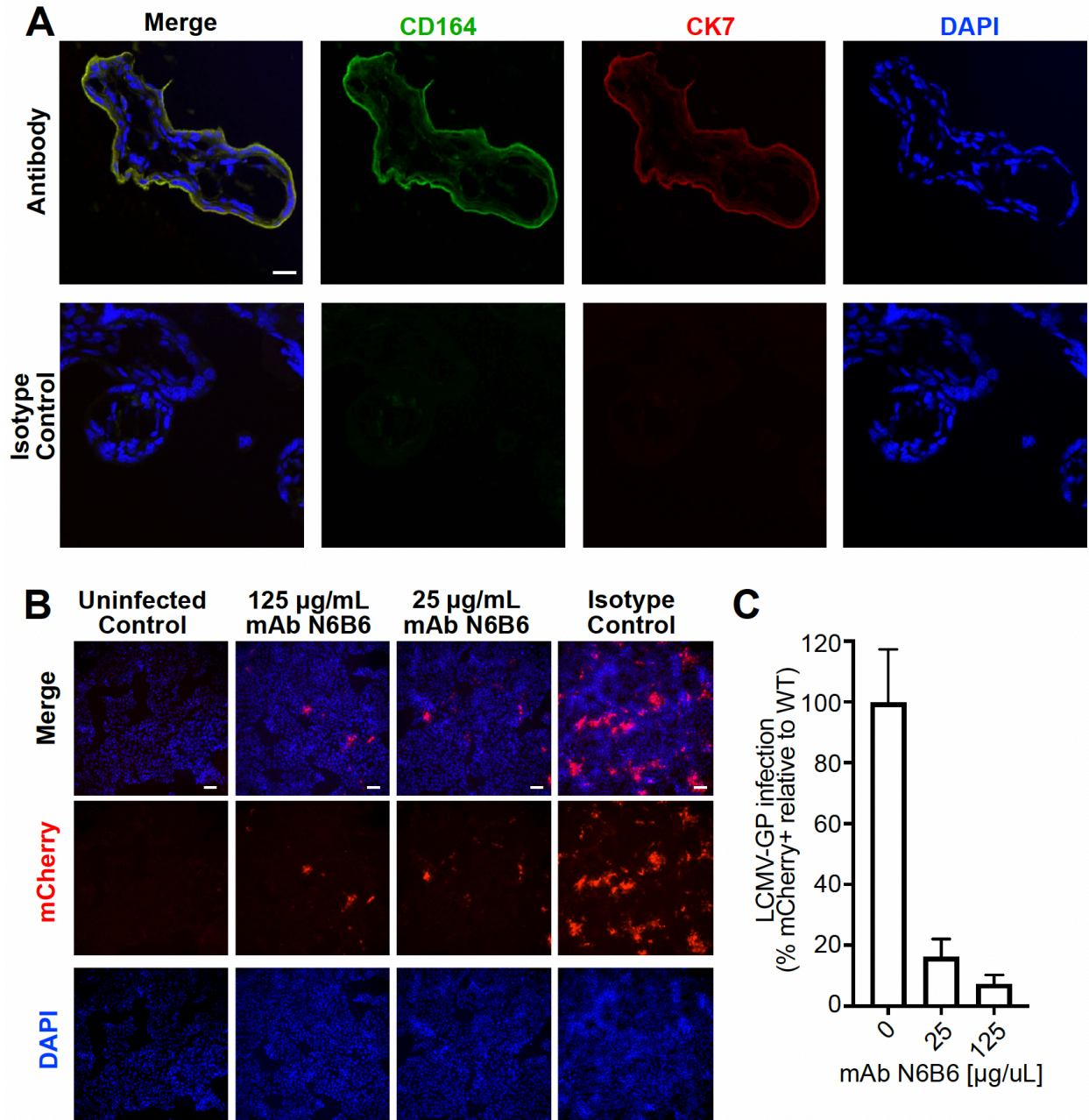


Figure 2.4 Characterization of CD164 as a therapeutic target in human placenta

(A) Double immunofluorescence staining for CD164 and CK7 or isotypes staining followed by counterstaining with DAPI in villous trophoblastic tissue. Original images were taken by confocal microscopy at 100x magnification. Scale bar represents 20 µm. (B) Immunofluorescence imaging of JEG-3 placenta cells pre-incubated with various concentrations of anti-CD164 mAb N6B6 and infected with r3LCMV-mCherry at MOI 0.5. Cells were fixed and imaged at 10x magnification 24 hpi. Scale bar represents 20 µm. (C) Quantification of percent infection of JEG-3 placenta cells pre-incubated with various concentrations of anti-CD164 mAb N6B6 and infected with r3LCMV-mCherry at MOI 0.5. Analysis was done on 4 FOV in 2 independent infections and normalized to infection control.

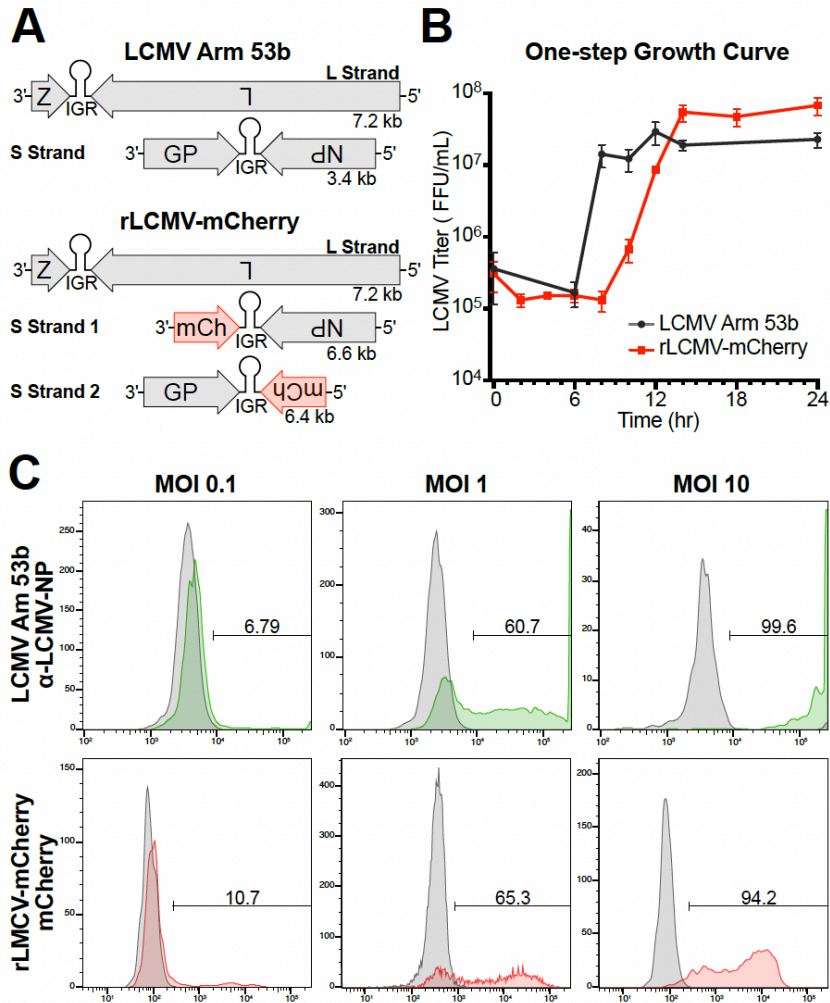


Figure 2.S1 Validation of recombinant virus rLCMV-mCherry infectivity

(A) Schematic representation of LCMV Arm 53b and rLCMV-mCherry genomes. (B) One-step growth curves of wildtype LCMV Arm 53b strain (black) and rLCMV-mCherry made in Arm 53b background (red) as measured by TCID₅₀ over a 24-hour time course. Error bars indicate standard error of three independent experiments. (C) Infection percentage of A549 cells infected at multiplicity of infection (MOI) 0.1, 1, and 10 with wildtype LCMV Arm 53b (top) or r3LCMV-mCherry (bottom) as measured at 24 hours post infection (hpi) using flow cytometry. Cells infected with Arm 53b were stained with anti-LCMV-NP monoclonal antibody (mAb) 113 primary and Alexa 488 secondary and measured for FITC signal in comparison with an uninfected control. Cells infected with rLCMV-mCherry were measured for PE-CF594 (mCherry) signal in comparison with an uninfected control.

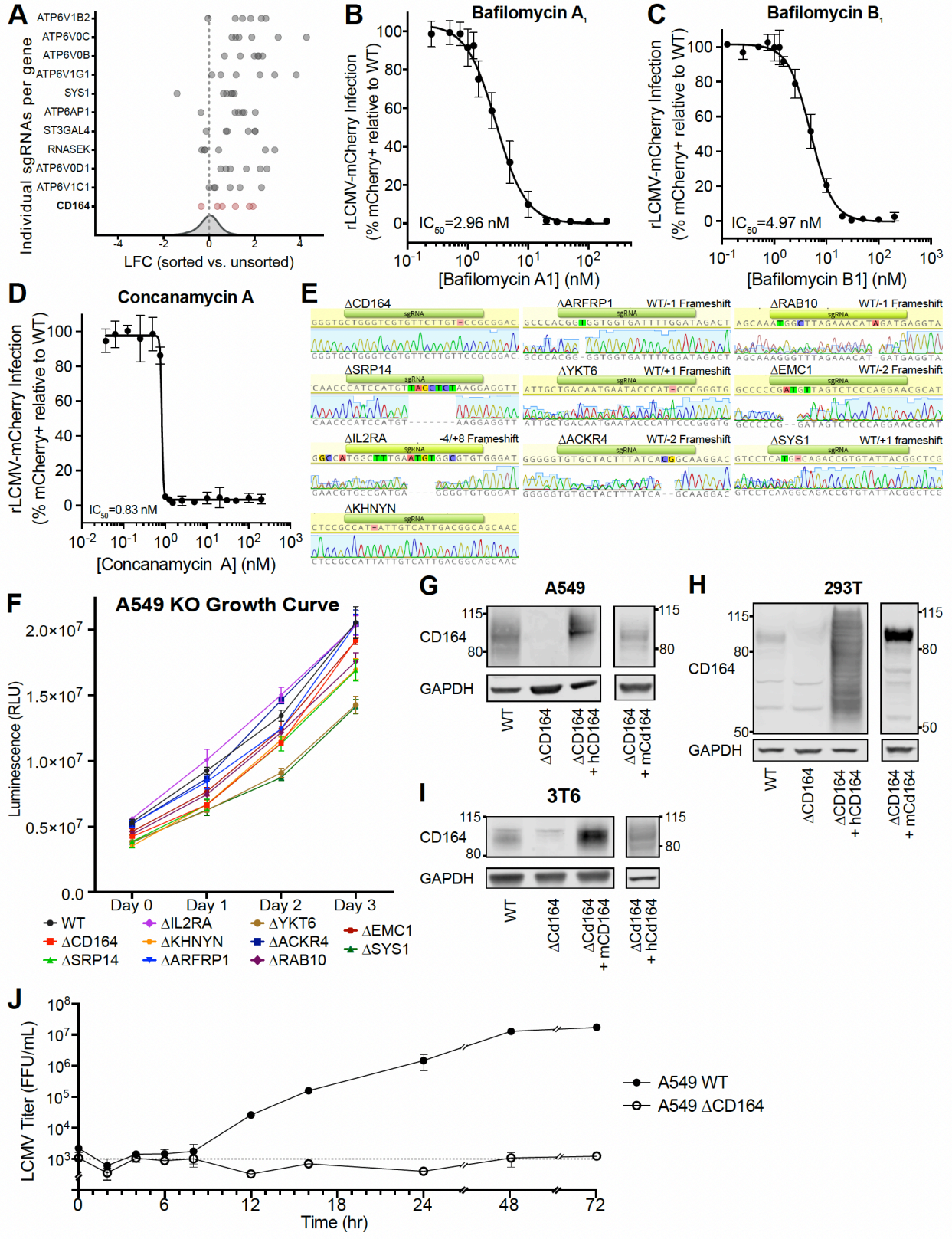


Figure 2.S2 Additional hit validation and characterization of gene-edited cells

(A) Log fold changes (LFC) of individual sgRNA of the top 10 scoring genes and CD164 (red) when comparing the infected and sorted cell population versus the uninfected cell population. Overall sgRNA distribution is shown at the bottom of the graph and dotted line indicates mean LFC of all sgRNAs. (B-D) Dose-response curve of v-ATPase inhibitors on rLCMV-mCherry infection at MOI 1 in A549 cells at 24 hpi, yielding (B) Bafilomycin A₁ IC₅₀ = 2.96 nM, (C) Bafilomycin B₁ IC₅₀ = 4.97 nM, and (D) Concanamycin A IC₅₀ = 0.83 nM. Error bars indicate standard error of three independent experiments. (E) Genotyping of clonal A549 where the target loci were PCR-amplified, Sanger-sequenced, and aligned to WT reference sequence. (F) Analysis of cell proliferation of WT and clonal A549 KO cells. Cells were plated in 96-well and proliferation was measured daily using Cell Titer Glo. Error bars indicate standard error from three separate well per cell line per time point. (G-I) Western blot analysis of WT, $\Delta CD164$, $\Delta CD164 + hCD164$, and $\Delta CD164 + mCd164$ for A549, 293T, and 3T6 cell lines. Human cell lines (A549 and 293T) were probed with anti-hCD164 antibody except for the *mCd164* addback which was probed with anti-mCd164 antibody. Mouse cell line 3T6 was probed with anti-mCd164 antibody except for the *hCD164* addback, which was probed with anti-hCD164 antibody. GAPDH was used as loading control. (J) One-step growth curves of rLCMV-Cherry on A549 WT or A549 $\Delta CD164$ cells as measured by TCID₅₀ over a 72-hour time course. Error bars indicate standard error of two independent experiments.

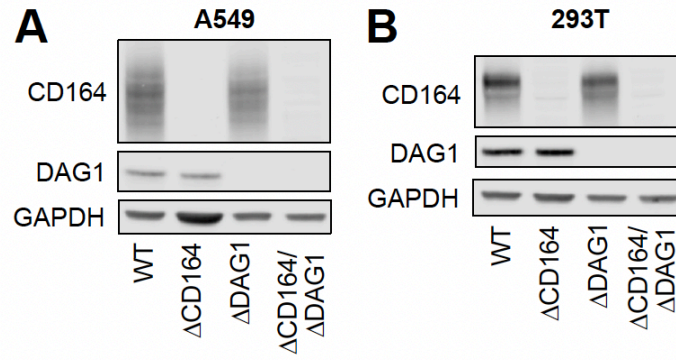


Figure 2.S3 Characterization of Δ CD164, Δ DAG1, and Δ CD164/ Δ DAG1 double KO cells

(A-B) Western blot analysis of Δ CD164, Δ DAG1, and Δ CD164/ Δ DAG1 double KO cells in (A) A549 or (B) 293T cell backgrounds. GAPDH was used as a loading control.

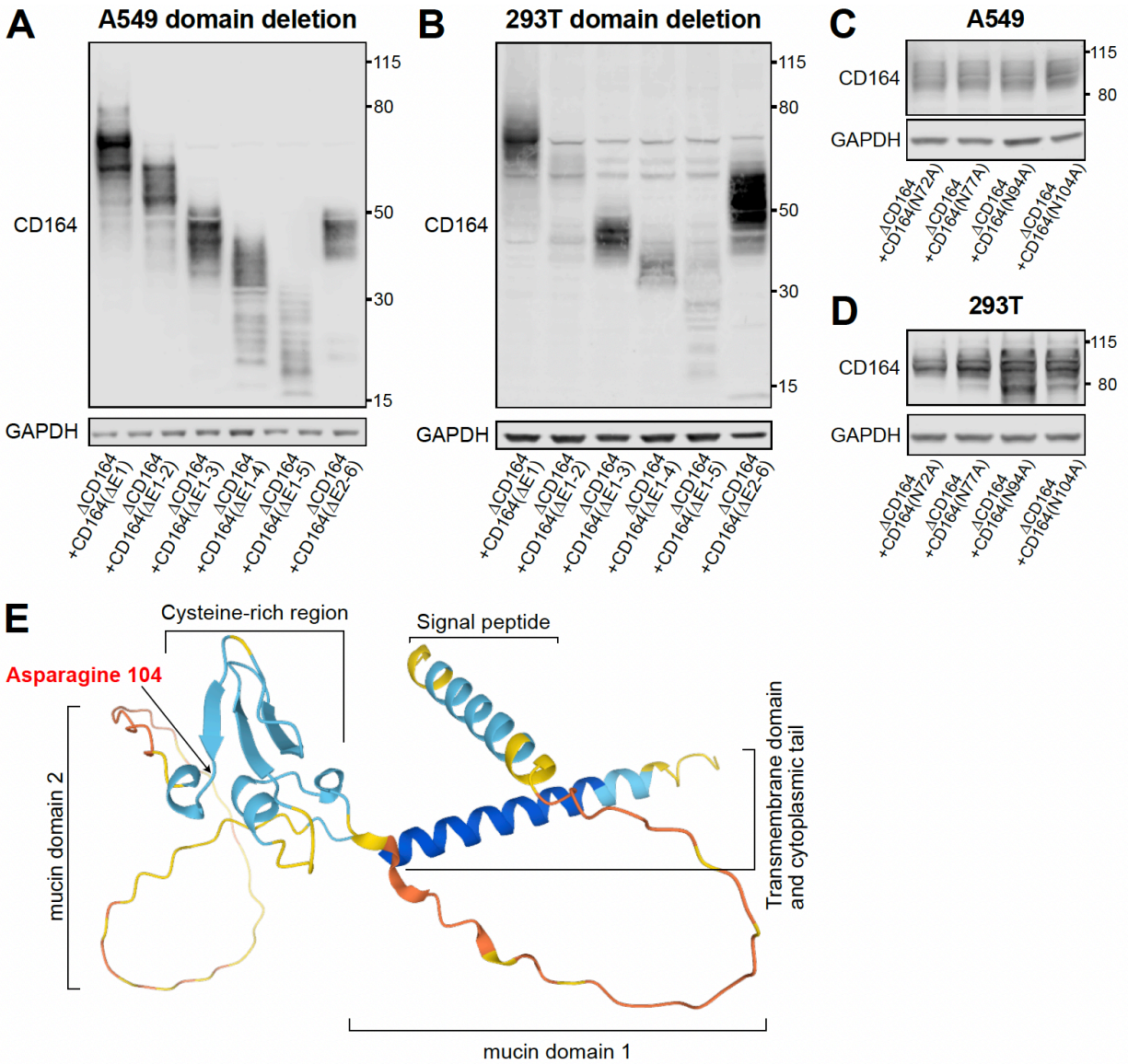


Figure 2.S4 Characterization of CD164 domain deletion and alanine mutagenesis add backs. (A-B) Western blot analysis of deletion domain addbacks in (A) A549 or (B) 293T cell backgrounds. (C-D) Western blot analysis of alanine mutagenesis addbacks in (C) A549 or (D) 293T cell backgrounds. All CD164 addbacks were probed with anti-FLAG antibody. GAPDH was used as a loading control. (E) AlphaFold prediction of CD164 protein structure. Prediction had low position error for the signal peptide, the cysteine-rich region, the transmembrane domain, and the cytoplasmic tail and high position error for the two mucin domains. Location of residue 104 is noted with an arrow.

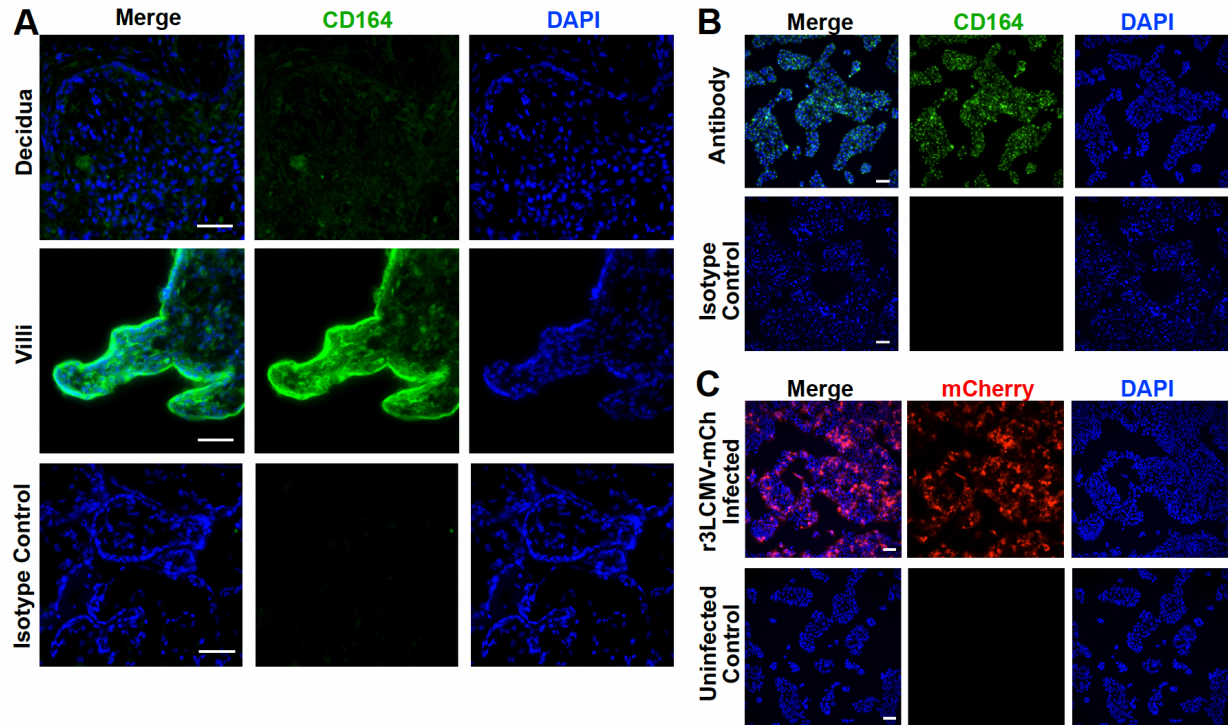


Figure 2.S5 Characterization of CD164 in placenta tissue and cell line

(A) Immunofluorescence staining of CD164 or isotype control followed by counterstaining with DAPI on placenta tissue at the maternal decidua and fetal villi. Original images taken at 40x magnification. Scale bar represents 50 μm . (B) Immunofluorescence imaging of CD164 or isotype control followed by counterstaining with DAPI on JEG-3 placenta cell line. Original images taken at 10x magnification. Scale bar represents 100 μm . (C) Immunofluorescence imaging JEG-3 placenta cell line with and without infection by r3LCMV-mCherry at MOI 1 and imaged at 24 hpi. Original images taken at 10x magnification. Scale bar represents 100 μm .

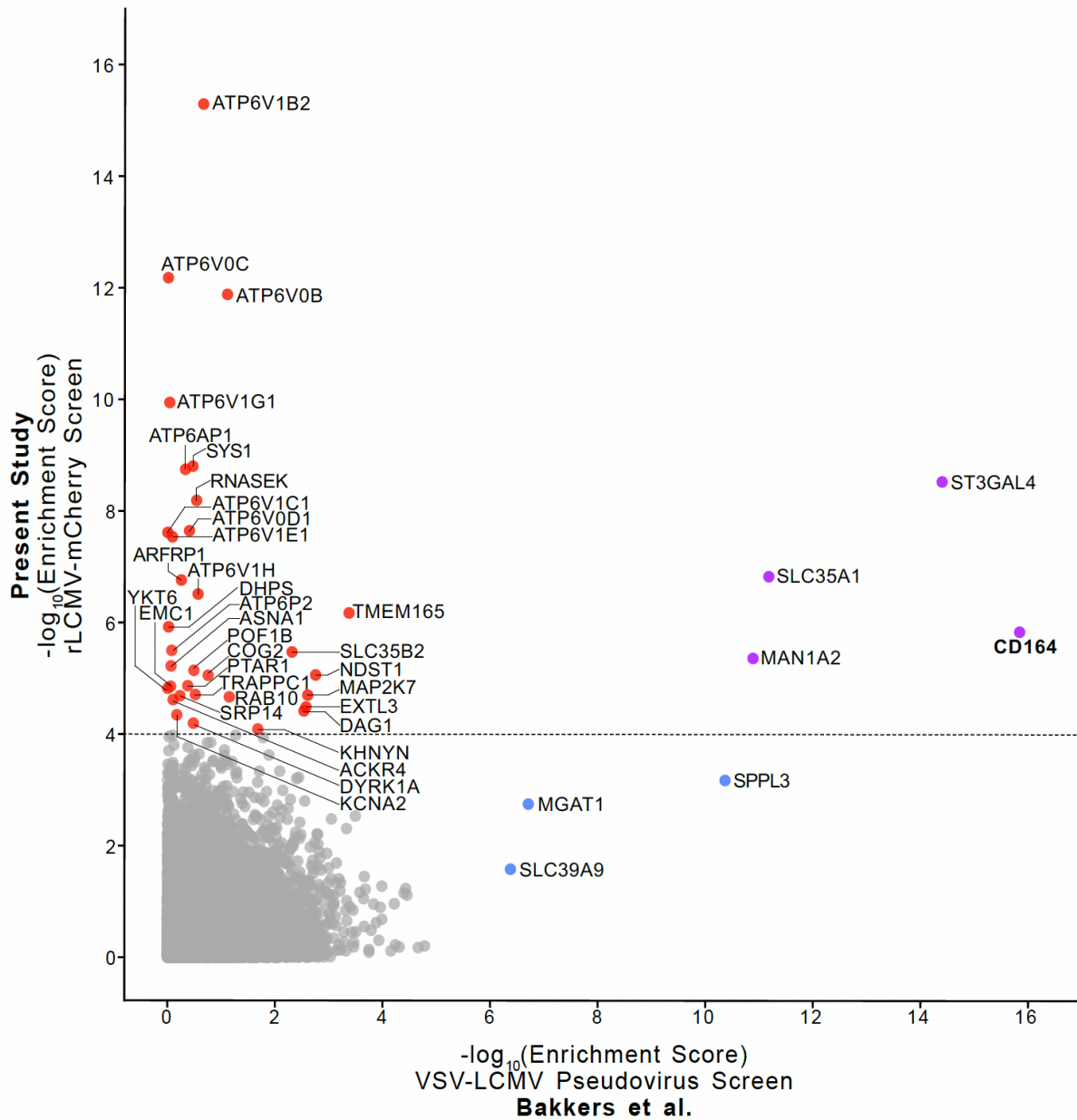


Figure 2.S6 Gene enrichment score comparison between two whole-genome CRISPR screens for LCMV host factors

The results of the whole-genome CRISPR screen of this study using the GeCKO v2 sgRNA library and rLCMV-mCherry compared to the results of Bakkers et al. (57) using the Brunello sgRNA library and VSV-LCMV pseudovirus. Enrichment scores are determined by MaGECK analysis. Genes enriched exclusively in this rLCMV-mCherry screen are colored in red. Genes enriched exclusively in the VSV-LCMV screen (SPPL3, MGAT1 and SLC39A9) are colored in blue. Genes enriched in both screens (ST3GAL4, SLC35A1, CD164, and MAN1A2) are colored in purple. All genes and their enrichment scores for the rLCMV-mCherry screen can be found in **Table S1**.

2.9 TABLES

Table 2.S1 rLCMV-mCherry whole genome CRISPR screen results

Table of the top genes identified by the rLCMV-mCherry whole genome CRISPR screen, along with the MAGeCK enrichment scores, p-values, false discovery rate, rank position, number of relevant sgRNAs, log fold change, and each ranking position of the gene's respective sgRNAs. (Top 90 rows reproduced).

id	posl score	posl p-value	posl fdr	posl rank	pos lgood	posl lfc	sgRNA rank
ATP6V1B2	5.08E-16	2.28E-07	0.00045	1	5	1.6504	[7, 8, 36, 60, 74, 110329]
ATP6V0C	6.59E-13	2.28E-07	0.00045	2	6	1.5977	[0, 16, 136, 327, 381, 1195]
ATP6V0B	1.31E-12	2.28E-07	0.00045	3	6	2.0716	[13, 14, 24, 65, 645, 1434]
ATP6V1G1	1.14E-10	2.28E-07	0.00045	4	6	1.7135	[4, 19, 21, 322, 14445, 88281]
SYS1	1.57E-09	2.28E-07	0.00045	5	5	0.87154	[479, 850, 1147, 1571, 1573, 29526]
ATP6AP1	1.79E-09	2.28E-07	0.00045	6	5	1.3773	[25, 44, 303, 581, 1619, 33101]
ST3GAL4	2.93E-09	2.28E-07	0.00045	7	5	1.2648	[3, 29, 63, 6482, 9780, 84532]
RNASEK	6.46E-09	2.28E-07	0.00045	8	3	0.12	[1, 2, 27794, 62478, 75859, 83468]
ATP6V0D1	2.26E-08	2.28E-07	0.00045	9	6	1.2889	[28, 114, 128, 1533, 23917, 50273]
ATP6V1C1	2.42E-08	2.28E-07	0.00045	10	5	0.59781	[22, 62, 132, 41868, 80450, 120785]
ATP6V1E1	2.91E-08	2.28E-07	0.00045	11	5	0.89351	[9, 32, 141, 2326, 22900, 31455]
SLC35A1	1.47E-07	1.60E-06	0.002888	12	6	1.1601	[165, 178, 243, 4121, 7384, 32308]
ARFRP1	1.73E-07	2.05E-06	0.003427	13	5	1.1707	[10, 113, 256, 5033, 8005, 93061]
ATP6V1H	3.08E-07	2.51E-06	0.00389	14	6	0.89869	[23, 41, 391, 1557, 18574, 30986]
TMEM165	6.68E-07	5.25E-06	0.007591	15	5	0.91781	[40, 203, 916, 1925, 56269, 118354]
DHPS	1.19E-06	7.99E-06	0.010775	16	6	0.8407	[397, 455, 1171, 2252, 9622, 71043]
CD164	1.44E-06	8.44E-06	0.010775	17	5	0.86904	[5, 164, 522, 4751, 29434, 42402]
ATP6AP2	3.15E-06	2.35E-05	0.027879	18	5	1.0132	[56, 422, 7538, 7702, 8994, 101241]
SLC35B2	3.35E-06	2.44E-05	0.027879	19	4	0.87476	[521, 941, 1069, 2994, 46011, 76899]
MAN1A2	4.30E-06	2.94E-05	0.031931	20	4	0.3847	[20, 64, 19091, 50807, 113733, 118367]
ASNA1	5.98E-06	4.36E-05	0.045026	21	5	0.79718	[58, 606, 848, 5949, 69308, 91361]
POF1B	7.19E-06	5.09E-05	0.04972	22	5	0.52915	[402, 4586, 4916, 10605, 11079, 13279]
NDST1	8.64E-06	6.18E-05	0.053663	23	3	0.44459	[55, 92, 4218, 83415, 95807, 103911]
COG2	8.81E-06	6.18E-05	0.053663	24	4	0.50572	[475, 514, 969, 28993, 75459, 104172]
hsa-mir-126	1.32E-05	5.27E-05	0.04972	25	3	0.9174	[561, 1316, 1977, 97316]
PTAR1	1.35E-05	9.29E-05	0.076458	26	6	0.70217	[142, 994, 1129, 22269, 33061, 34479]
EMC1	1.39E-05	9.51E-05	0.076458	27	4	0.77239	[33, 264, 1138, 31001, 96523, 109659]
YKT6	1.50E-05	9.88E-05	0.076556	28	4	0.81855	[46, 597, 2723, 4613, 33792, 97623]
TRAPPC1	1.95E-05	0.00012	0.086714	29	4	0.81559	[619, 1717, 2188, 4993, 67239, 115641]
MAP2K7	1.99E-05	0.000122	0.086714	30	6	0.72639	[759, 955, 2403, 5030, 23604, 81239]
SRP14	2.05E-05	0.000124	0.086714	31	6	0.84409	[283, 506, 3225, 5084, 20244, 39895]
RAB10	2.14E-05	0.000131	0.088645	32	5	0.81688	[123, 1305, 2783, 5162, 48679, 95984]
ACKR4	2.40E-05	0.000146	0.09616	33	3	0.49042	[517, 823, 1382, 93311, 102632, 120409]
EXTL3	3.25E-05	0.000198	0.126529	34	6	0.59031	[878, 1025, 8193, 14958, 20375, 38648]
DAG1	3.81E-05	0.000227	0.141018	35	6	0.53501	[2198, 6332, 8139, 13819, 17012, 81930]
KCNA2	4.50E-05	0.00026	0.156766	36	5	0.52118	[2456, 4015, 4192, 6412, 86234, 88761]
DYRK1A	6.35E-05	0.000341	0.199759	37	5	0.68406	[697, 5324, 14063, 14326, 19478, 111789]
KHNYN	8.08E-05	0.000431	0.245831	38	4	0.61317	[792, 1714, 2144, 20630, 90063, 91673]
CRTC1	0.000105	0.00057	0.311015	39	6	0.79221	[2100, 3432, 6759, 17347, 33142, 48691]
SLC35G2	0.000105	0.000573	0.311015	40	5	0.69557	[2155, 3809, 6952, 10727, 22250, 119697]
POLR2B	0.000111	0.000608	0.322024	41	4	0.66208	[2525, 3127, 4556, 8470, 42329, 90406]
IL2RA	0.000115	0.000627	0.324022	42	6	0.49454	[3431, 4833, 9263, 17412, 22750, 50469]

id	posl score	posl p-value	posl fdr	posl rank	pos lgood	posl fcr	sgRNA rank
BBS2	0.000141	0.000763	0.382876	43	6	0.55437	[2285, 14100, 21387, 21822, 48722, 51603]
RNF4	0.000143	0.000776	0.382876	44	3	0.37169	[71, 388, 9031, 17048, 91684, 106647]
NT5C1A	0.000157	0.000856	0.412651	45	5	0.45673	[2222, 12100, 15311, 17725, 20178, 24654]
DUX2	0.000167	0.000906	0.427142	46	3	0.15883	[133, 418, 15341, 25509, 31643, 82224]
POM121L2	0.000171	0.000933	0.430904	47	4	0.96554	[1275, 3451, 7297, 9731, 35796, 112982]
VPS51	0.000192	0.001048	0.473907	48	5	0.82818	[472, 5614, 17813, 17938, 25949, 66518]
TRABD	0.000198	0.001076	0.476359	49	6	0.45804	[1838, 8577, 26385, 45397, 54544, 55119]
BOD1L2	0.000226	0.001213	0.526436	50	6	0.33173	[15923, 18666, 19127, 42132, 55480, 56601]
RS1	0.000235	0.001262	0.529037	51	5	0.59492	[210, 5559, 8011, 17121, 27343, 108622]
SNX32	0.000236	0.001268	0.529037	52	6	0.45113	[14443, 14458, 36721, 37875, 40353, 57069]
AGRP	0.000241	0.001294	0.529703	53	4	0.5245	[6128, 7088, 8167, 10846, 65923, 106936]
FAM134B	0.000266	0.001428	0.563186	54	6	0.42412	[1736, 4210, 30956, 51121, 54795, 58342]
CCDC140	0.000294	0.001576	0.584081	55	5	0.4472	[4114, 5317, 11055, 24558, 28922, 93947]
C15orf61	0.000294	0.001579	0.584081	56	5	0.70576	[5272, 5919, 8391, 12293, 14499, 28944]
NTN4	0.000298	0.001591	0.584081	57	6	0.38539	[3841, 28724, 42525, 47994, 56938, 59618]
ZNF91	0.00031	0.001657	0.584081	58	3	0.28013	[481, 571, 10284, 95550, 99835, 107035]
ATP12A	0.00031	0.001659	0.584081	59	4	0.45908	[2375, 9133, 9271, 11737, 39632, 46272]
CPEB4	0.000313	0.001673	0.584081	60	6	0.64296	[783, 2194, 3551, 13427, 48500, 81512]
IQGAP1	0.00032	0.001704	0.584081	61	6	0.42267	[2146, 11294, 16598, 23736, 48913, 60462]
ATP6V1F	0.000324	0.001723	0.584081	62	1	-0.23909	[6, 13717, 48601, 67555, 80054, 99419]
LAP3	0.000338	0.001794	0.598858	63	5	0.82229	[1565, 3351, 10171, 12060, 31153, 41007]
POP4	0.000345	0.001833	0.602535	64	4	0.64355	[250, 611, 18299, 37377, 93691, 94842]
TPTE	0.000352	0.001872	0.606103	65	3	0.07	[121, 617, 35192, 52040, 63099, 111950]
WDR78	0.000365	0.001931	0.608337	66	5	0.71024	[1601, 1613, 7052, 12372, 37947, 81348]
hsa-mir-4520a	0.000365	0.001403	0.563186	67	3	0.36952	[205, 996, 57966, 84011]
CGNL1	0.000366	0.001935	0.608337	68	4	0.4499	[270, 1066, 3780, 61005, 91078, 116218]
ATP6V1A	0.000367	0.001708	0.584081	69	3	0.39431	[575, 769, 24691, 85866, 99060, 110295]
CYLD	0.000372	0.00197	0.610537	70	4	0.73509	[306, 936, 8569, 12454, 109568, 117327]
HIST3H2A	0.00038	0.002016	0.616023	71	3	0.30711	[97, 642, 10938, 28450, 67998, 114600]
RPS5	0.000394	0.002085	0.628232	72	3	0.14107	[183, 654, 39019, 44397, 54620, 97985]
OXER1	0.000404	0.002131	0.629169	73	5	0.95943	[221, 791, 3944, 16495, 44269, 65634]
HIST1H2BK	0.000414	0.002191	0.629169	74	2	-0.11035	[258, 670, 57664, 80591, 88296, 116321]
PTPLA	0.000417	0.002204	0.629169	75	4	0.33136	[392, 672, 23694, 55434, 102097, 107784]
EGFL8	0.000429	0.002258	0.636299	76	2	0.12361	[87, 681, 77739, 96814, 100214, 115634]
RGP1	0.000437	0.002295	0.638297	77	6	0.90601	[768, 2118, 9404, 16884, 32023, 67362]
GPX8	0.000459	0.002397	0.651918	78	5	0.64754	[3420, 5403, 6552, 13338, 81508, 98223]
CC2D1A	0.00046	0.002404	0.651918	79	5	0.83119	[738, 4598, 7739, 13351, 67548, 68641]
LYRM7	0.00047	0.002173	0.629169	80	5	0.54888	[610, 10066, 12938, 18849, 60726, 86166]
DEF8	0.000491	0.002557	0.670677	81	4	0.56061	[3177, 6664, 12871, 13641, 99849, 113566]
COG1	0.0005	0.002605	0.670677	82	5	0.60946	[180, 735, 8200, 36697, 46130, 66730]
MOV10	0.000505	0.002633	0.670677	83	6	0.45456	[13664, 28480, 30551, 35834, 53750, 65914]
SLC37A2	0.000507	0.00264	0.670677	84	4	0.43996	[550, 1827, 10979, 13798, 105156, 119637]
SMARCB1	0.00051	0.002658	0.670677	85	6	0.53447	[510, 2203, 5574, 25811, 48737, 66015]
OR2B3	0.000528	0.002741	0.68368	86	5	0.29403	[11427, 15389, 16208, 28885, 33605, 107889]
TAAR2	0.000543	0.002814	0.686228	87	3	0.30019	[537, 766, 20696, 28871, 76049, 105735]
PCOLCE2	0.000543	0.002815	0.686228	88	4	0.43586	[1350, 1763, 4402, 30114, 63457, 68581]
PAPL	0.000551	0.002858	0.688944	89	4	0.6785	[3047, 4216, 4743, 14179, 30063, 103692]

Table 2.S2 Gene ontology overrepresentation analysis

Results of using the PANTHER Overrepresentation test to analyze the top hits identified through the rLCMV-mCherry whole genome CRISPR screen. Fold enrichment, raw p-value, false discovery rate, as well as enriched gene members were identified for the gene ontologies biological process, molecular function, and cellular component.

Analysis Type:	PANTHER Overrepresentation Test (Released 20210224)					
Annotation Version and Release Date:	GO Ontology database DOI: 10.5281/zenodo.5228828 Released 2021-08-18					
Reference List:	Homo sapiens (all genes in database)					
Test Type:	FISHER					
Correction:	FDR					
	Homo sapiens - REFLIST (20595)	Input: LCMV Screen Hits (394)	Input (expected) over/under	fold Enrichment raw p-value FDR	Gene_Members	
GO biological process complete						
proton transmembrane transport (GO:1902600)	134	14	2.56 +	5.46	8.08E-07	1.27E-02 P6V1F/ATP55
GO molecular function complete						
7S RNA binding (GO:0008312)	7	4	0.13 +	29.87	3.64E-05	1.77E-02 SRP14/SRP68/SRP19/SRP72
proton-transporting ATPase activity, rotational mechanism (GO:0046961)	17	9	0.33 +	27.67	6.23E-10	ATP6V1E1/ATP6V0D1/ATP6V1B2/ATP6V0C/ATP6V1A/ATP6V1G
ATPase-coupled cation transmembrane transporter activity (GO:0019829)	47	11	0.9 +	12.23	9.16E-09	KCNJ8/ATP6V1E1/ATP6V0D1/ATP6V1B2/ATP6V0C/ATP6V1A/A
ATPase-coupled transmembrane transporter activity (GO:0042626)	111	11	2.12 +	5.18	1.93E-05	8.94E-06 TP6V1G1/ATP12A/ATP6V1H/ATP6V1C1/ATP6V1F
ATPase activity, coupled to transmembrane movement of ions, rotational mechanism	17	9	0.33 +	27.67	6.23E-10	KCNJ8/ATP6V1E1/ATP6V0D1/ATP6V1B2/ATP6V0C/ATP6V1A/A
ATPase-coupled ion transmembrane transporter activity (GO:0042625)	17	9	0.33 +	27.67	6.23E-10	ATP6V1E1/ATP6V0D1/ATP6V1B2/ATP6V0C/ATP6V1A/ATP6V1G
pyrophosphate hydrolysis-driven proton transmembrane transporter activity	20	10	0.38 +	26.14	1.04E-10	ATP6V1E1/ATP6V0D1/ATP6V1B2/ATP6V0C/ATP6V1A/ATP6V1G
proton transmembrane transporter activity (GO:0015078)	129	13	2.47 +	5.27	2.88E-06	1/ATP6V08/ATP12A/ATP6V1H/ATP6V1C1/CLCN4/ATP6V1F/ATP

GO cellular component complete	Homo sapiens - REFLIST (20595)	Input: LCMV Screen Hits (394)	Input (expected)	over/ under	fold Enrichment	raw P-value	FDR	Gene_Members
vacuolar proton-transporting V-type ATPase, V1 domain (GO:0000221)	8	6	0.15 +		39.2	1.11E-07	3.69E-05 1	ATP6V1E1/ATP6V1B2/ATP6V1A/ATP6V1G1/ATP6V1H/ATP6V1C
vacuolar proton-transporting V-type ATPase complex (GO:0016471)	20	9	0.38 +		23.52	1.90E-09		ATP6V1E1/ATP6V0D1/ATP6AP2/ATP6V1B2/ATP6V1A/ATP6V1G 1/ATP6V1H/ATP6V1C1/ATP6V1F
proton-transporting V-type ATPase complex (GO:0033176)	27	12	0.52 +		23.23	4.01E-12		ATP6V1E1/ATP6V0D1/ATP6AP2/ATP6V1B2/ATP6AP1/ATP6V0C /ATP6V1A/ATP6V1G1/ATP6V1H/ATP6V1C1/ATP6V1F
proton-transporting two-sector ATPase complex (GO:0016469)	50	13	0.96 +		13.59	1.32E-10		ATP6V1E1/ATP6V1G1/ATP6V0B/ATP6V1H/ATP6V1C1/ATP6V0C /ATP6V1A/ATP6V1G1/ATP6V0B/ATP6V1H/ATP6V1C1/ATP6V1F /ATP6V1A/ATP6V1G1/ATP6V1H/ATP6V1C1/ATP6V1F
proton-transporting V-type ATPase, V1 domain (GO:0033180)	11	8	0.21 +		38.02	9.06E-10		ATP6V1E1/ATP6V1B2/ATP6AP1/ATP6V1A/ATP6V1G1/ATP6V1H /ATP6V1C1/ATP6V1F
proton-transporting two-sector ATPase complex, catalytic domain (GO:0033178)	19	8	0.36 +		22.01	2.33E-08		ATP6V1E1/ATP6V1B2/ATP6AP1/ATP6V1A/ATP6V1G1/ATP6V1H /ATP6V1C1/ATP6V1F
signal recognition particle, endoplasmic	6	4	0.11 +		34.85	2.35E-05		9.27E-06 /ATP6V1C1/ATP6V1F
signal recognition particle (GO:0048500)	6	4	0.11 +		34.85	2.35E-05		5.18E-03 SRP14/SRP68/SRP19/SRP72
vesicle tethering complex (GO:0099023)	63	7	1.21 +		5.81	3.38E-04		5.83E-03 SRP14/SRP68/SRP19/SRP72
trans-Golgi network membrane (GO:0032588)	101	9	1.93 +		4.66	2.31E-04		4.79E-02 COG4/COG2/TRAPP1/VPS51/EXOC3/TRAPP4/COG1 COG4/COG2/RGP1/SYS1/ARFRP1/VPS51/TMEM165/RAB6A/CO 4.59E-02 G1
lysosomal membrane (GO:0005765)	397	19	7.59 +		2.5	3.60E-04		ATP6V1E1/ATP6V0D1/ATP6AP2/TMEM30A/ATP6V1B2/ATP6V0 C/GPLD1/ATP6V1A/CD164/ATP6V1G1/ATP6V0B/TMEM165/PS 4.48E-02 EN1/ATP6V1H/ATP6V1C1/CLCN4/TPCNI/HSP90A81/OCLN
lytic vacuole membrane (GO:0098852)	397	19	7.59 +		2.5	3.60E-04		ATP6V1E1/ATP6V0D1/ATP6AP2/TMEM30A/ATP6V1B2/ATP6V0 C/GPLD1/ATP6V1A/CD164/ATP6V1G1/ATP6V0B/TMEM165/PS 4.77E-02 EN1/ATP6V1H/ATP6V1C1/CLCN4/TPCNI/HSP90A81/OCLN

Table 2.S3 Antibodies used in this study

Antibodies and their respective dilutions that were used in this study are listed along with their vendor and catalog numbers.

Name	Vendor	Dilution	Cat No
mouse anti-human CD164 clone N6B6	BD Pharmingen	1:200	551296
mouse IgG2a, κ Isotype Control	BD Pharmingen	1:200	555571
anti-VSV-G clone 8G5F11	Kerafast	1:10 000	EB0010
mouse anti-LCMV NP 1.1.3	Gift from of Michael J. Buchmeier	1:200	N/A
rat anti-human Cytokeratin-7 Clone 7D3	Sigma Millipore	1:50	MABT1490
goat anti-mouse IgG Alexa Fluor 488	Thermo Fisher Scientific	1:1000	A-11029
Goat Anti-Rat IgG H&L (TRITC)	abcam	1:1000	ab7094
donkey anti-rabbit 680	LI-COR	1:15 000	926-68073
donkey anti-mouse 800	LI-COR	1:15 000	926-32212
goat anti-mouse Cd164 polyclonal	R&D Systems	1:200	AF3118
rabbit anti-human CD164 polyclonal	Thermo Fisher Scientific	1:200	PA5-80418
rabbit anti-FLAG M2	Cell Signaling Technologies	1:500	14793S
mouse anti-GAPDH	Santa Cruz Biotechnology	1:500	sc-32233
NucBlue			

Table 2.S4 Oligonucleotides used in this study

This table lists all oligonucleotides used in this study for conducting the whole genome CRISPR screen, for producing sgRNA and genotyping resulting knockout cell lines, for cloning cDNA of interest, and for cloning mammarenavirus glycoprotein-pseudotyped VSV of interest.

A. CRISPR screen primers

Primer name	Primer sequence
GeCKO PCR1 F	ttgcatatacgatacaaggctggt
GeCKO PCR1 R	gatgaataactgccatttgtctcaa
GeCKO PCR2 F	AATGATACGGCGACCACCGAGATCTACACNNNNNNNNNACACTCTTTCCCTACACGA CGCTCTTCCGATCTtcttggtaaaggacgaaacacc
GeCKO PCR2 R	CAAGCAGAAGACGGCATAACGAGATNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCT CTTCCGATCTCGGACTAGCCTTATTTTAACTTGCTATTTTC
GeCKO custom sequencing primer	GCTCTTCCGATCTTCTTGTGGAAAGGACGAAACACCG

Where “NNNNNNNN” indicates index sequences as follows:

Index ID	Index	Index2 ID	Index2
D701	ATTACTCG	D501	AGGCTATA
D702	TCCGGAGA	D502	GCCTCTAT
D703	CGCTCATT	D503	AGGATAGG
D704	GAGATTCC	D504	TCAGAGCC

B. sgRNA and primers

sgRNA sequences for generating KO

Gene	sgRNA	oligo F	oligo R	Genotyping primer F	Genotyping primer R
CD164	CTGGGTCGTTCCTTGCCG	CACCGTGGGTCGTTCCTTGCCG	AAACCGGACAAGAACACACGCCCCAGC	CCCCAAGTGAGGCTAGGAAT	TTGCAGAAGCTCCCAGTTTTT
SRP14	CATCCATGTTAGCTCTAAGG	CACCGCATCCATGTTAGCTCTAAGG	AAACCCCTTAGAGCTAACATGGATGC	CTACTGTGGGGAAACCAGAA	CCTTCTCAGGGCAAGAGTCA
IL2RA	TGGCTTTGAATGTGGCGTGT	CACCGTGGCTTTGAATGTGGCGTGT	AAACACACGCCCCACATTCAAAGCCAC	TCCAGCCCTAGGCCAACACAGAT	GTGCTTTGGAGAGGAGCTTG
KHNYN	TGCCGTCAATGACAATATGG	CACCGTGCCGTCAATGACAATATGG	AAACCCCATATTGTCATTCACGGCAC	CCTGGGGAAGGAGGAGATAG	GCCTATCCTGGGACACTGAA
ARFRP1	ATCCAAAATCACACCACCG	CACCGATCCAAAATCACACCACCG	AAACCGGTGGTGTGATTTTGGATC	CTAAGGGTTTCAGCTGGGACA	TCAGCTGCACGTTTCACTACC
YKT6	TGACAAATGAATACCCATCCC	CACCGTGACAAATGAATACCCATCCC	AAACGGGATGGGTATTCATTTGTCAC	GCATATTGAGCTTCCGAGTGG	TGTGGTGCAACTGACTGACA
RAB10	ATGGCTTAGAAACATAGATG	CACCGATGGCTTAGAAACATAGATG	AAACCATCTATGTTTCTTAAGCCATC	TCCCTTTCAGATTTTTCATCCA	TTTGGAGACGGAGTCTTGCT
ACKR4	TGTGCTACTTTATFCACAGCA	CACCGTGTGCTACTTTTATFCACAGCA	AAACTGCTGTGATAAAGTAGCACAC	TGCCATCTTGTGTGACATAC	ACGGGCTCTGAGATTTGAGT
EMC1	TTCCCTGGGAGACTAACATCG	CACCGTTCCTGGGAGACTAACATCG	AAACCCGATGTTAGTCTCCAGGAAC	GTTGCAGTGAAGCCAAAGATCA	GGAAGGAGACTGGGGACTC
SYS1	CGTAATACACGGTCTGCATG	CACCGGTAATACACGGTCTGCATG	AAACCATGCAGACCGGTGTATTACCG	GCTGGAGCTTTTGCCTCTCTA	ACATCCAGGCCCTTTTCTTTT
mouse					
Cd164	CAAGTCAGTCCGGTTCACCT	CACCGCAAGTCAGTCCGGTTCACCT	AAACAGGTGAACCCGCACTGACTTGC	TCCTTGTGCAATCCATCCCTT	GAAGTAGAGCAGGAGGACC

C. cDNA cloning primers

Primer name	primer
pLenti-mouse-Cd164-EcoRV-F	tgtggtggaatttctgcagatgccaccATGTCGGGCTCTCCCGCCG
pLenti-human-CD164-EcoRV-F	tgtggtggaatttctgcagatgccaccATGTCGGGCTCTCCCGCTCACT
pLenti-Origene-EcoRV-R	gggaaaaagatcctgtttctcTCATTTATCGTCATCGTCTTTGTAGTCAGATCC GCCACCATCAAAGGTCGACTTCCGCTC

Domain deletion mutagenesis	primer
DDM_CD164_delta-tail_rev	CGGCCGCCACTGTGCTGGATTTACTTTATCGTCGTCATCCTTGTAATCAAGAAA GAAAATTACAGCCTGC
DDM_CD164_SSP_rev	CGCGGACAGCACGCAGAGCA
DDM_CD164_SSP-E2_fwd	GCGTGCTGTCCGCGGAAACCTGTGAAGGTCGAAACAGC
DDM_CD164_SSP-E3_fwd	GCGTGCTGTCCGCGGATGAGAGCTATTGTTACATAACTCAACAG
DDM_CD164_SSP-E4_fwd	GCGTGCTGTCCGCGGTTTCCACGGCCACTCCAG
DDM_CD164_SSP-E5_fwd	GCGTGCTGTCCGCGGCTAAACCCACAGTTCAGCCC
DDM_CD164_SSP-E6_fwd	GCGTGCTGTCCGCGGGTACAACAAATAACACTGTGACTCC
DDM_CD164_E1_rev	GAAACTGGCTGCATCTTCTGGTGCCGGAGTGG
DDM_CD164_E1-E6_fwd	GATGCAGCCAGTTTCATTGGAGG

Site-directed mutagenesis	primer
SDM_CD164_N72A_fwd	CCTGTTTTGCTGTTAGCGTTGTTAATACTACC
SDM_CD164_N72A_rev	GGTAGTAGCAACAACGCTAACATTAAAACAG
SDM_CD164_N77A_fwd	GCGTTGTTGCTACTACCTGCTTTTGGATAG
SDM_CD164_N77A_rev	CAAAAGCAGGTAGTAGCAACAACGCTAAC
SDM_CD164_N94A_fwd	GTTACATGCCTCAACAGTTAGTGATTG
SDM_CD164_N94A_rev	CTGTTGAGGCATGTGAACAATAGCTCTC
SDM_CD164_N104A_fwd	CAAGTGGGGCCACGACAGACTTCTG
SDM_CD164_N104A_rev	CTGTCGTGGCCCCACTTGACAATCAC

D. Cloning pseudotyped VSV primers

primer name	primer
LCMV-WE-HPI_pCAGGS_fwd	gctgtctcatcattttggcaaag
LCMV-WE-HPI_pCAGGS_rev	gggaaaagatcctgttctcTCATTTATCGTCATCGTCTTTGTAGTC AGATCCGCCACCgcgctctttccagatagtttttac
GTOV-Flag_pCAGGS_fwd	gtctcatcattttggcaaagGCCACCATGGGTCAATTGTTCA
GTOV-Flag_pCAGGS_rev	gggaaaagatcctgttctcTCATTTATCGTCATCGTCTTTGTAGTC AGATCCGCCACCgTCGTGGTTCTTGTGCCAGG
LCMV-Arm4_F260L_fwd	CTAAGTTCCTCACTAGGAGACTAGCGGGCAC
LCMV-Arm4_F260L_rev	CTCCTAGTGAGGAACTTAGTCTTCTCTTGGGAAAGGAG
LCMV-Arm4_S153F_fwd	GAGGGAAC TTCAACTATAAGGCAGTATCCTGC
LCMV-Arm4_S153F_rev	CTTATAGTTGAAGTTCCCTCTGATACTGAGGTGTAGG
LCMV-WE-HPI_S133T_fwd	acaaaaagtattttgaccatacactcatgagtatag
LCMV-WE-HPI_S133T_rev	ggtcaaaatactttttgttgaaagcggagg
LCMV-WE-HPI_T211A_fwd	ggctgggcaggttcagatggcaagaccac
LCMV-WE-HPI_T211A_rev	ctgaacctgcccagccccagccacttctc
LCMV-WE-HPI_V94A_fwd	gtgctcagccaacaactctcatcactacatcagtatgg
LCMV-WE-HPI_V94A_rev	gagagttggttgctgagcacgcattgggcatc
LCMV-WE-HPI_Y155H_fwd	ggaattccaaccacaaagcagtgctcttgatthtaac
LCMV-WE-HPI_Y155H_rev	cactgctttgtggttggaattccctctgatac
LCMV-WE-HPI_S153F_Y155H_fwd	cagaggggaatttcaaccacaaagcagtgctcttgatthtaac
LCMV-WE-HPI_S153F_Y155H_rev	cactgctttgtggttgaaattccctctgatactgagggtg

2.10 REFERENCES

1. Radoshitzky SR, Buchmeier MJ, Charrel RN, Clegg JCS, Gonzalez J-PJ, Günther S, Hepojoki J, Kuhn JH, Lukashevich IS, Romanowski V, Salvato MS, Sironi M, Stenglein MD, de la Torre JC, Ictv Report Consortium null. 2019. ICTV Virus Taxonomy Profile: Arenaviridae. *J Gen Virol* 100:1200–1201.
2. Stenglein MD, Sanchez-Migallon Guzman D, Garcia VE, Layton ML, Hoon-Hanks LL, Boback SM, Keel MK, Drazenovich T, Hawkins MG, DeRisi JL. 2017. Differential Disease Susceptibilities in Experimentally Reptarenavirus-Infected Boa Constrictors and Ball Pythons. *J Virol* 91:e00451-17.
3. Hepojoki J, Hepojoki S, Smura T, Szirovicza L, Dervas E, Prähauser B, Nufer L, Schraner EM, Vapalahti O, Kipar A, Hetzel U. 2018. Characterization of Haartman Institute snake virus-1 (HISV-1) and HISV-like viruses-The representatives of genus Hartmanivirus, family Arenaviridae. *PLoS Pathog* 14:e1007415.
4. Shi M, Lin X-D, Chen X, Tian J-H, Chen L-J, Li K, Wang W, Eden J-S, Shen J-J, Liu L, Holmes EC, Zhang Y-Z. 2018. The evolutionary history of vertebrate RNA viruses. *Nature* 556:197–202.
5. Burri DJ, da Palma JR, Seidah NG, Zanotti G, Cendron L, Pasquato A, Kunz S. 2013. Differential recognition of Old World and New World arenavirus envelope glycoproteins by subtilisin kexin isozyme 1 (SKI-1)/site 1 protease (S1P). *J Virol* 87:6406–6414.
6. Traub E. 1935. A FILTERABLE VIRUS RECOVERED FROM WHITE MICE. *Science* 81:298–299.
7. Childs JE, Glass GE, Korch GW, Ksiazek TG, Leduc JW. 1992. Lymphocytic choriomeningitis virus infection and house mouse (*Mus musculus*) distribution in urban Baltimore. *Am J Trop Med Hyg* 47:27–34.

8. Centers for Disease Control and Prevention (CDC). 2008. Brief report: Lymphocytic choriomeningitis virus transmitted through solid organ transplantation--Massachusetts, 2008. *MMWR Morb Mortal Wkly Rep* 57:799–801.
9. Wright R, Johnson D, Neumann M, Ksiazek TG, Rollin P, Keech RV, Bonthius DJ, Hitchon P, Grose CF, Bell WE, Bale JF. 1997. Congenital lymphocytic choriomeningitis virus syndrome: a disease that mimics congenital toxoplasmosis or Cytomegalovirus infection. *Pediatrics* 100:E9.
10. Barton LL, Mets MB, Beauchamp CL. 2002. Lymphocytic choriomeningitis virus: emerging fetal teratogen. *Am J Obstet Gynecol* 187:1715–1716.
11. Meyer BJ, de la Torre JC, Southern PJ. 2002. Arenaviruses: genomic RNAs, transcription, and replication. *Curr Top Microbiol Immunol* 262:139–157.
12. Beyer WR, Pöpplau D, Garten W, von Laer D, Lenz O. 2003. Endoproteolytic processing of the lymphocytic choriomeningitis virus glycoprotein by the subtilase SKI-1/S1P. *J Virol* 77:2866–2872.
13. Borrow P, Oldstone MB. 1992. Characterization of lymphocytic choriomeningitis virus-binding protein(s): a candidate cellular receptor for the virus. *J Virol* 66:7270–7281.
14. Eschli B, Quirin K, Wepf A, Weber J, Zinkernagel R, Hengartner H. 2006. Identification of an N-terminal trimeric coiled-coil core within arenavirus glycoprotein 2 permits assignment to class I viral fusion proteins. *J Virol* 80:5897–5907.
15. Di Simone C, Zandonatti MA, Buchmeier MJ. 1994. Acidic pH triggers LCMV membrane fusion activity and conformational change in the glycoprotein spike. *Virology* 198:455–465.
16. Burns JW, Buchmeier MJ. 1991. Protein-protein interactions in lymphocytic choriomeningitis virus. *Virology* 183:620–629.

17. Cao W, Henry MD, Borrow P, Yamada H, Elder JH, Ravkov EV, Nichol ST, Compans RW, Campbell KP, Oldstone MB. 1998. Identification of alpha-dystroglycan as a receptor for lymphocytic choriomeningitis virus and Lassa fever virus. *Science* 282:2079–2081.
18. Spiropoulou CF, Kunz S, Rollin PE, Campbell KP, Oldstone MBA. 2002. New World arenavirus clade C, but not clade A and B viruses, utilizes alpha-dystroglycan as its major receptor. *J Virol* 76:5140–5146.
19. Holt KH, Crosbie RH, Venzke DP, Campbell KP. 2000. Biosynthesis of dystroglycan: processing of a precursor propeptide. *FEBS Lett* 468:79–83.
20. Kunz S, Rojek JM, Kanagawa M, Spiropoulou CF, Barresi R, Campbell KP, Oldstone MBA. 2005. Posttranslational modification of alpha-dystroglycan, the cellular receptor for arenaviruses, by the glycosyltransferase LARGE is critical for virus binding. *J Virol* 79:14282–14296.
21. Rojek JM, Spiropoulou CF, Campbell KP, Kunz S. 2007. Old World and clade C New World arenaviruses mimic the molecular mechanism of receptor recognition used by alpha-dystroglycan's host-derived ligands. *J Virol* 81:5685–5695.
22. Smelt SC, Borrow P, Kunz S, Cao W, Tishon A, Lewicki H, Campbell KP, Oldstone MB. 2001. Differences in affinity of binding of lymphocytic choriomeningitis virus strains to the cellular receptor alpha-dystroglycan correlate with viral tropism and disease kinetics. *J Virol* 75:448–457.
23. Kunz S, Sevilla N, Rojek JM, Oldstone MBA. 2004. Use of alternative receptors different than alpha-dystroglycan by selected isolates of lymphocytic choriomeningitis virus. *Virology* 325:432–445.
24. Shimojima M, Kawaoka Y. 2012. Cell surface molecules involved in infection mediated by lymphocytic choriomeningitis virus glycoprotein. *J Vet Med Sci* 74:1363–1366.

25. Volland A, Lohmüller M, Heilmann E, Kimpel J, Herzog S, von Laer D. 2021. Heparan sulfate proteoglycans serve as alternative receptors for low affinity LCMV variants. *PLoS Pathog* 17:e1009996.
26. Brouillette RB, Phillips EK, Patel R, Mahauad-Fernandez W, Moller-Tank S, Rogers KJ, Dillard JA, Cooney AL, Martinez-Sobrido L, Okeoma C, Maury W. 2018. TIM-1 Mediates Dystroglycan-Independent Entry of Lassa Virus. *J Virol* 92:e00093-18.
27. Jae LT, Raaben M, Herbert AS, Kuehne AI, Wirchnianski AS, Soh TK, Stubbs SH, Janssen H, Damme M, Saftig P, Whelan SP, Dye JM, Brummelkamp TR. 2014. Virus entry. Lassa virus entry requires a trigger-induced receptor switch. *Science* 344:1506–1510.
28. Raaben M, Jae LT, Herbert AS, Kuehne AI, Stubbs SH, Chou Y-Y, Blomen VA, Kirchhausen T, Dye JM, Brummelkamp TR, Whelan SP. 2017. NRP2 and CD63 Are Host Factors for Lassa Virus Cell Entry. *Cell Host Microbe* 22:688-696.e5.
29. Sanjana NE, Shalem O, Zhang F. 2014. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* 11:783–784.
30. Emonet SF, Garidou L, McGavern DB, de la Torre JC. 2009. Generation of recombinant lymphocytic choriomeningitis viruses with trisegmented genomes stably expressing two additional genes of interest. *Proc Natl Acad Sci U S A* 106:3473–3478.
31. Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, Irizarry RA, Liu JS, Brown M, Liu XS. 2014. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* 15:554.
32. Jae LT, Raaben M, Riemersma M, van Beusekom E, Blomen VA, Velds A, Kerkhoven RM, Carette JE, Topaloglu H, Meinecke P, Wessels MW, Lefeber DJ, Whelan SP, van Bokhoven H, Brummelkamp TR. 2013. Deciphering the glycosylome of dystroglycanopathies using haploid screens for lassa virus entry. *Science* 340:479–483.

33. Shin H-W, Kobayashi H, Kitamura M, Waguri S, Suganuma T, Uchiyama Y, Nakayama K. 2005. Roles of ARFRP1 (ADP-ribosylation factor-related protein 1) in post-Golgi membrane trafficking. *J Cell Sci* 118:4039–4048.
34. Zhang T, Hong W. 2001. Ykt6 forms a SNARE complex with syntaxin 5, GS28, and Bet1 and participates in a late stage in endoplasmic reticulum-Golgi transport. *J Biol Chem* 276:27480–27487.
35. Khoury G, Lee MY, Ramarathinam SH, McMahon J, Purcell AW, Sonza S, Lewin SR, Purcell DFJ. 2021. The RNA-Binding Proteins SRP14 and HMGB3 Control HIV-1 Tat mRNA Processing and Translation During HIV-1 Latency. *Front Genet* 12:680725.
36. Kisaka JK, Ratner L, Kyei GB. 2020. The Dual-Specificity Kinase DYRK1A Modulates the Levels of Cyclin L2 To Control HIV Replication in Macrophages. *J Virol* 94:e01583-19.
37. Tsunetsugu-Yokota Y, Honda M. 1990. Effect of cytokines on HIV release and IL-2 receptor alpha expression in monocytic cell lines. *J Acquir Immune Defic Syndr (1988)* 3:511–516.
38. Mi H, Ebert D, Muruganujan A, Mills C, Albu L-P, Mushayamaha T, Thomas PD. 2021. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res* 49:D394–D403.
39. Forgac M. 2007. Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology. *Nat Rev Mol Cell Biol* 8:917–929.
40. Dröse S, Bindseil KU, Bowman EJ, Siebers A, Zeeck A, Altendorf K. 1993. Inhibitory effect of modified bafilomycins and concanamycins on P- and V-type adenosinetriphosphatases. *Biochemistry* 32:3902–3906.
41. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigartyo CA-K, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist P-H, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M,

- Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Pontén F. 2015. Proteomics. Tissue-based map of the human proteome. *Science* 347:1260419.
42. Watt SM, Bühring HJ, Rappold I, Chan JY, Lee-Prudhoe J, Jones T, Zannettino AC, Simmons PJ, Doyonnas R, Sheer D, Butler LH. 1998. CD164, a novel sialomucin on CD34(+) and erythroid subsets, is located on human chromosome 6q21. *Blood* 92:849–866.
43. Zannettino AC, Bühring HJ, Niutta S, Watt SM, Benton MA, Simmons PJ. 1998. The sialomucin CD164 (MGC-24v) is an adhesive glycoprotein expressed by human hematopoietic progenitors and bone marrow stromal cells that serves as a potent negative regulator of hematopoiesis. *Blood* 92:2613–2628.
44. Havens AM, Jung Y, Sun YX, Wang J, Shah RB, Bühring HJ, Pienta KJ, Taichman RS. 2006. The role of sialomucin CD164 (MGC-24v or endolyn) in prostate cancer metastasis. *BMC Cancer* 6:195.
45. Whitt MA. 2010. Generation of VSV pseudotypes using recombinant Δ G-VSV for studies on virus entry, identification of entry inhibitors, and immune responses to vaccines. *J Virol Methods* 169:365–374.
46. Fukushi S, Tani H, Yoshikawa T, Saijo M, Morikawa S. 2012. Serological assays based on recombinant viral proteins for the diagnosis of arenavirus hemorrhagic fevers. *Viruses* 4:2097–2114.
47. Chan JY, Lee-Prudhoe JE, Jorgensen B, Ihrke G, Doyonnas R, Zannettino AC, Buckle VJ, Ward CJ, Simmons PJ, Watt SM. 2001. Relationship between novel isoforms, functionally important domains, and subcellular distribution of CD164/endolyn. *J Biol Chem* 276:2139–2152.
48. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Židek A, Green T, Tunyasuvunakool K, Petersen S, Jumper J, Clancy E, Green R, Vora A, Lutfi M, Figurnov M, Cowie A, Hobbs N, Kohli P, Kleywegt G, Birney E,

- Hassabis D, Velankar S. 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 50:D439–D444.
49. Doyonnas R, Yi-Hsin Chan J, Butler LH, Rappold I, Lee-Prudhoe JE, Zannettino AC, Simmons PJ, Bühring HJ, Levesque JP, Watt SM. 2000. CD164 monoclonal antibodies that block hemopoietic progenitor cell adhesion and proliferation interact with the first mucin domain of the CD164 receptor. *J Immunol* 165:840–851.
50. Jamieson DJ, Kourtis AP, Bell M, Rasmussen SA. 2006. Lymphocytic choriomeningitis virus: an emerging obstetric pathogen? *Am J Obstet Gynecol* 194:1532–1536.
51. Bonthius DJ. 2012. Lymphocytic choriomeningitis virus: an underrecognized cause of neurologic disease in the fetus, child, and adult. *Semin Pediatr Neurol* 19:89–95.
52. Barton LL, Mets MB. 2001. Congenital lymphocytic choriomeningitis virus infection: decade of rediscovery. *Clin Infect Dis* 33:370–374.
53. Genbacev O, Larocque N, Ona K, Prakobphol A, Garrido-Gomez T, Kapidzic M, Bárcena A, Gormley M, Fisher SJ. 2016. Integrin $\alpha 4$ -positive human trophoblast progenitors: functional characterization and transcriptional regulation. *Hum Reprod* 31:1300–1314.
54. Fisher S, Genbacev O, Maidji E, Pereira L. 2000. Human cytomegalovirus infection of placental cytotrophoblasts in vitro and in utero: implications for transmission and pathogenesis. *J Virol* 74:6808–6820.
55. Rizzuto G, Brooks JF, Tuomivaara ST, McIntyre TI, Ma S, Rideaux D, Zikherman J, Fisher SJ, Erlebacher A. 2022. Establishment of fetomaternal tolerance through glycan-mediated B cell suppression. *Nature* 603:497–502.
56. Dammann O, Leviton A. 1997. Maternal intrauterine infection, cytokines, and brain damage in the preterm newborn. *Pediatr Res* 42:1–8.

57. Bakkers MJG, Moon-Walker A, Herlo R, Brusica V, Stubbs SH, Hastie KM, Sapphire EO, Kirchhausen TL, Whelan SPJ. 2022. CD164 is a host factor for lymphocytic choriomeningitis virus entry. *Proc Natl Acad Sci U S A* 119:e2119676119.
58. Buchmeier MJ, Lewicki HA, Tomori O, Oldstone MB. 1981. Monoclonal antibodies to lymphocytic choriomeningitis and pichinde viruses: generation, characterization, and cross-reactivity with other arenaviruses. *Virology* 113:73–85.
59. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, Virgin HW, Listgarten J, Root DE. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* 34:184–191.
60. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8:2281–2308.
61. Campeau E, Ruhl VE, Rodier F, Smith CL, Rahmberg BL, Fuss JO, Campisi J, Yaswen P, Cooper PK, Kaufman PD. 2009. A versatile viral system for expression and depletion of proteins in mammalian cells. *PLoS One* 4:e6529.
62. Sena-Esteves M, Tebbets JC, Steffens S, Crombleholme T, Flake AW. 2004. Optimized large-scale production of high titer lentivirus vector pseudotypes. *J Virol Methods* 122:131–139.

3 IDENTIFICATION OF HOST FACTORS INVOLVED IN LCMV VIRAL EGRESS AND FUTURE DIRECTIONS

3.1 BACKGROUND

Viruses are obligate intracellular pathogens that must enter host cells and hijack host pathways in order to propagate. Elucidation of host factors important for productive infection and understanding the transcriptional changes caused by viral infection can lead to identification of therapeutic targets of viral infection. From RNA interference to CRISPR/Cas9, whole-genome screens focused on understanding the host determinants of viral infections such as HIV, influenza, Zika, and Dengue have all led to the discovery of a myriad of unique genes that play instrumental roles in host-pathogen interactions [1–5]. These high throughput methods are limited in scope, however, as they generally are only able to assay the entry, uncoating, and replication steps of the viral life cycle [6,7]. The final stage of viral reproduction is assembly and budding, together often referred to as viral egress, is known to be notoriously challenging to study. This is because to understand the host factors involved in viral egress, candidate-based approaches are taken, as no simple high-throughput method currently exist that is able to probe this life cycle stage.

Viral egress involves a process where chemically distinct macromolecules, ranging from large protein complexes to viral genome strands, are transported and gathered within a cell where they can be assembled into a viral particle [8]. These nascent virions are then released from the infected host and able to infect nearby cells repeat the replication cycle. To appropriately assess whether viral egress has successfully occurred, a secondary infection due to virions produced from the primary infected cell must be confirmed. Thus, to develop an assay to study viral egress, at a minimum, there must be a reporter system in place that is able to identify and differentiate primary from secondary infections.

There are additional requirements to establishing a high through-put assay for studying the egress portion of the viral life cycle. Traditional whole-genome functional genomics screens are conducted using bulk methods, making it impossible to determine which genetic perturbation has a disrupted production of virions. This is due to the presence of non-relevant genetic perturbations as well as non-targeting controls, neither of which inhibit viral egress in any way and will always allow the release of productive virions into the shared supernatant and cause secondary infections. Thus, to be able to conduct a whole-genome functional genomic screen to elucidate host factors involved in viral egress, a novel high-throughput assay must be developed such that individual perturbations can still be simultaneously assayed in isolation, each with their own secondary infection reporters.

Current methods for identifying novel host factors involved in the assembly and budding stage of the viral life cycle remains challenging. Candidate-based approaches to understanding the host-pathogen interactions that occur during the final stage of virus reproduction both limited and biased due to low throughput [9,10]. Whole-genome functional screens has changed the landscape for gene discovery and identification of novel therapeutic targets in viral infection. Should these techniques be applied to studying viral egress, a new type of viral therapeutics can be established that would, instead of disrupting current infections, be able to halt and inhibit

future spread of the disease. By preventing viral egress, these viral therapies would be able to directly impact the contagiousness of an infectious disease and play a dominant role in public health and safety.

3.2 EXPERIMENTAL APPROACHES

3.2.1 Engineered split reporter system allows for identification of successful viral egress

To address the first challenge in the development of a high-throughput functional genomics assay for investigating viral egress host factors, we needed to develop a reporter system that is able to identify successful viral egress. This was established using the split GFP system [11] while leveraging the flexible genome packaging ability of lymphocytic choriomeningitis virus (LCMV) [12,13]. Thus, to differentiate between cells with a primary infection (cells infected using titered virus) and cells with a secondary infection (cells infected by viable virions originating from primary infected cells) a recombinant tri-segmented lymphocytic choriomeningitis virus (rLCMV-Duo) was first recovered from plasmids. In contrast to wild-type LCMV, this rescued virus contains three genomic strands: the L strand and two S strands where each contains one of either NP or GP as well as the mCherry-3xGFP11 reporter of interest (**Figure 3.1A**).

Next, we made a HeLaS3 cell line that constitutively expresses GFP1-10 (HeLaS3 dC9K-GFP1-10) to use as a secondary infection reporter cell. One-step growth curves demonstrated similar growth kinetics for rLCMV-Duo in HeLaS3 dC9K cells and HeLaS3 dC9K GFP1-10 cells (**Figure 3.1B**) [14]. Furthermore, a 24-hour time course also demonstrated that the detection of the mCherry reporter and GFP reporter are comparable in timescales, allowing for accurate prediction of infection (**Figure 3.1C**). Altogether, the simultaneous use of the rLCMV-Duo reporter virus with the HeLaS3 dC9K GFP1-10 reporter

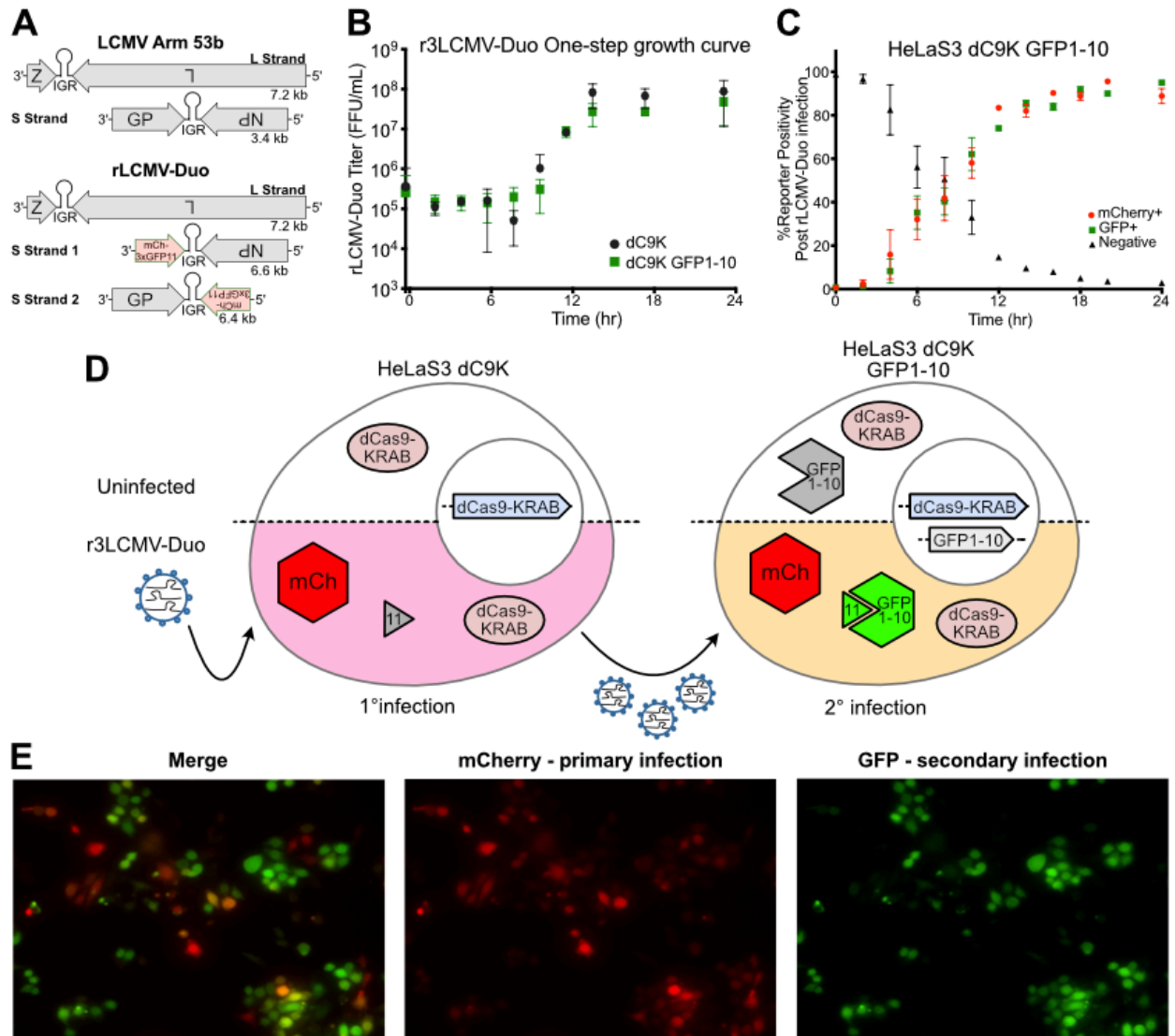


Figure 3.1 Recombinant LCMV containing mCherry and GFP11 reporters (rLCMV-Duo) allows for identification of successful viral egress.

(A) Schematic representation of LCMV Arm4 53b (wildtype) and rLCMV-Duo (recombinant). (B) One-step growth curves of rLCMV-Duo in either HeLaS3 dCas9-KRAB (dC9K) (black) or HeLaS3 dC9K-GFP1-10 (green) cells as measured by TCID₅₀ over a 24-h time course. Error bars indicate standard errors from three independent experiments. (C) Reporter positivity of mCherry (red), GFP (green), or negative (black) in HeLaS3 dC9K-GFP1-10 infected at multiplicity of infection (MOI) 10 as measured by flow cytometry over a 24-h time course. (D) Schematic demonstrating how use of rLCMV-Duo together with HeLaS3 dC9K and dC9K-GFP1-10 would be able to differentiate primary and secondary infections. (E) Representative microscopy images demonstrating that primary infection (mCherry positive only) and secondary infection (mCherry positive and GFP positive) can be differentiated in a bulk population following infection with rLCMV-Duo.

cell line allows for accurate differentiation of primary and secondary infections: Expression of mCherry indicates a successful primary infection, while GFP positivity due to the combination of the 3xGFP11 from rLCMV-Duo and the constitutively expressed GFP1-10 produced from the HeLaS3 dC9K-GFP1-10 cells would indicate a secondary infection (**Figure 3.1D**).

Indeed, when primary HeLaS3 dC9K cells were infected with rLCMV-Duo, mCherry was expressed. When these infected cells were co-cultured with uninfected HeLaS3 dC9K GFP1-10 cells, the HeLaS3 dC9K GFP1-10 cells began expressing both mCherry and GFP at 16 hours post infection (hpi) (**Figure 3.1.E**), thus allowing for differentiation between cells experiencing primary and secondary infections. As such, we have established a system in which we would be able to accurately assess whether viral egress has successfully occurred following genetic perturbations.

3.2.2 Droplet-based microfluidics allows independent assays to be conducted in a high-throughput manner

One of the major challenges of screening for host factors important in viral egress is attaining an appropriate throughput. Due to egress being a viral life stage that is notoriously challenging to assay, most attempts to explore the host-pathogen interactions during this stage rely on candidate-based approaches. We propose that by using droplet-based microfluidics, we would be able to simultaneously assay hundreds of thousands of egress events in parallel.

As previously described, simply applying current bulk methods for conducting genome-wide functional screens would not be able to elucidate host factors playing a role in viral assembly and budding. This is because in a bulk screen where all genetic perturbations are simultaneously assayed for successful secondary infection within one pot (**Figure 3.2**), so long as a single perturbation is able to generate viable nascent virions (for example, in non-targeting

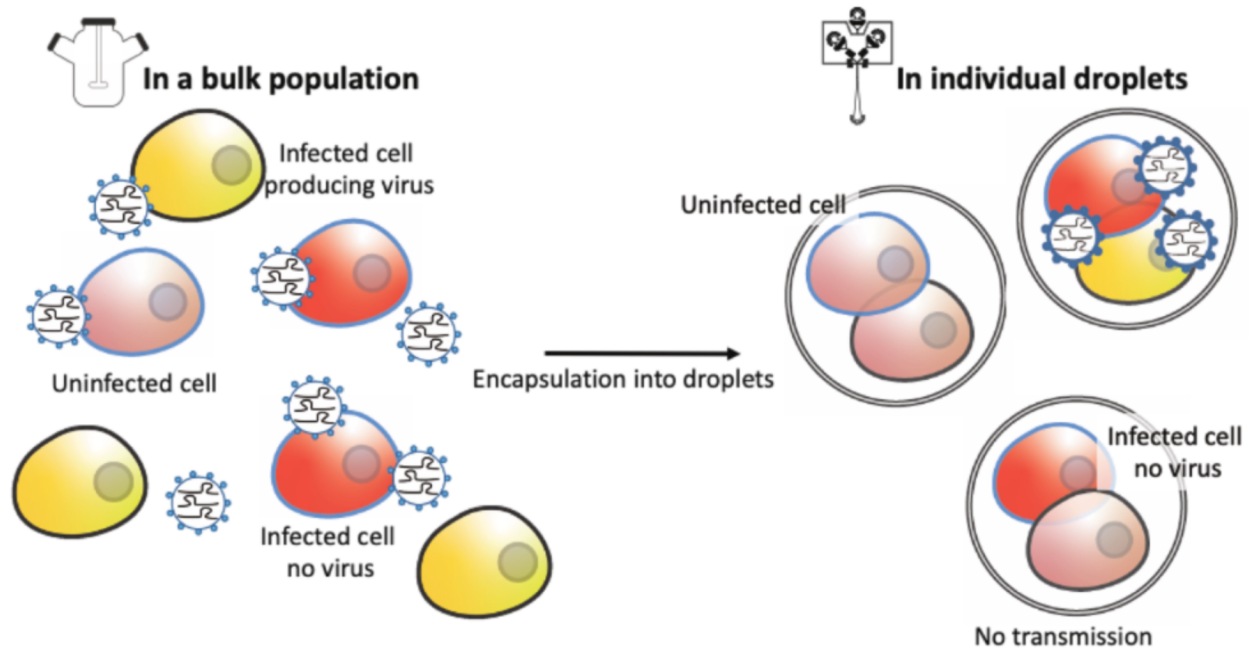


Figure 3.2 Schematic comparing phenotypes of infection outcomes in a bulk population and in individual droplets.

In a bulk population, all reporter cells (HeLaS3 dC9K GFP1-10) will be infected by rLCMV-mCh-GFP11 and become mCherry positive and GFP positive so long as one KD cell is able to produce virus. In the droplet setting, droplets containing KD cells that are uninfected (mCherry negative) will be co-encapsulated with a mCherry negative and GFP negative reporter cell. Droplets containing KD cells that are infected and producing viable viral progeny will be co-encapsulated with a mCherry positive and GFP positive reporter cell. A droplet that contains an infected KD cell (mCherry positive) but unable to produce viable viral progeny, however, will be co-encapsulated with an uninfected reporter cell that is mCherry negative and GFP negative. These different droplet phenotypes allow us to differentiate KD cells representing a gene that is 1) important for viral entry, 2) unimportant for viral infection, and, 3) important for viral egress respectively.

controls or non-relevant targets), the secondary infection reporters would express successful egress. Thus, it becomes impossible to retroactively identify the which genetic perturbations causes a lack of viable virions being produced. However, if we were able to separate each genetic perturbation such that they are all individually and independently assayed for productive viral egress, then a high-throughput screen would be possible. We believe one such method would be to co-encapsulate each primary infected cell that has a gene of interest knocked down with a secondary reporter cell.

Droplet-based microfluidics allow for encapsulation of cells, organic materials, and/or reagents into discrete, uniformly distributed, miniaturized reaction chambers [15]. By segmenting an aqueous flow with an immiscible carrier fluid, a natural barrier to diffusion is formed. This two-phase system exploits the pressure differential between each phase and the microfluidic geometry to produce monodispersed picoliter-sized droplets at rates up to 100 kHz [16]. Using this method, cell encapsulation into droplets allows individual cells to be assayed independently in an isolated microenvironment. Furthermore, fluorinated oil, which is able to dissolve much higher concentrations of O₂, can be used as the carrier sheath to allow encapsulated cells to remain viable for extended periods of time for downstream assays [16–18]. Single-cell assays can thus be conducted in a high-throughput manner using oil-water emulsion droplets [16,19].

In the context of assessing successful viral egress, by simply assessing the florescence of each droplet, perhaps using a flow cytometer, we would be able to readily identify cases where the genetic perturbation experienced by the primary infected cell either 1) causes no effect on viral egress and thus both mCherry-positive/GFP-negative cells and mCherry-positive/GFP-positive cells are present within the same droplet, 2) has an effect on viral entry and thus all cells in the droplet remain mCherry-negative/GFP-negative, or 3) has an effect on

viral egress and therefore the primary infected cell is mCherry-positive/GFP-negative but there are no GFP-positive cells within the same droplet.

We designed a microfluidic device that is able to co-encapsulate infected primary cells with secondary reporter cells into stable, monodispersed droplets (**Figure 3.3A**). While there exists many options to co-encapsulate two cell populations into a single droplet, we chose to iterate on a passive co-axial flow-focusing device prototyped using polydimethylsiloxane (PDMS) such that desired droplet size formation can be adjusted with minimal impact on device complexity [20,21]. The resulting droplets (**Figure 3.3B-D**) should be predominantly a water in oil emulsion, where the cells are able to survive for a minimum of 16 hours or until viral egress and expression of the GFP reporter has occurred.

As we require the co-encapsulation of two independent cell populations into a single droplet, cell loading can be approximated by a double Poisson distribution. It is imperative that the genetic perturbations we are testing in the primary infected cells are loaded into droplets at a rate of one cell per droplet. This requires encapsulation of primary cells to occur at a distribution of $\lambda = 0.1$, such that most droplets (90%) will not capture a primary infected cell, few (9%) of droplets will contain one primary infected cell, but only a very small minority (>1%) will contain two or more primary infected cells [16]. However, if we were to follow this stringent rule for a double Poisson distribution, fewer than 1% of droplets we produced would contain a single primary infected cell together with a single secondary reporter cell. This would mean that to achieve the necessary representation of guides in a whole-genome functional screen, we would need to process over 100x more cells than required in a bulk screen. This would render such a screen infeasible due to space and time requirements alone [10,16].

Thus, we will utilize a suspension cell type that can survive prolonged incubation at extremely high seeding concentrations. By using HeLaS3 cells, which are able to survive at high concentrate ions in droplets over 32 hours (**Figure 3.3E**) when in the presence of density

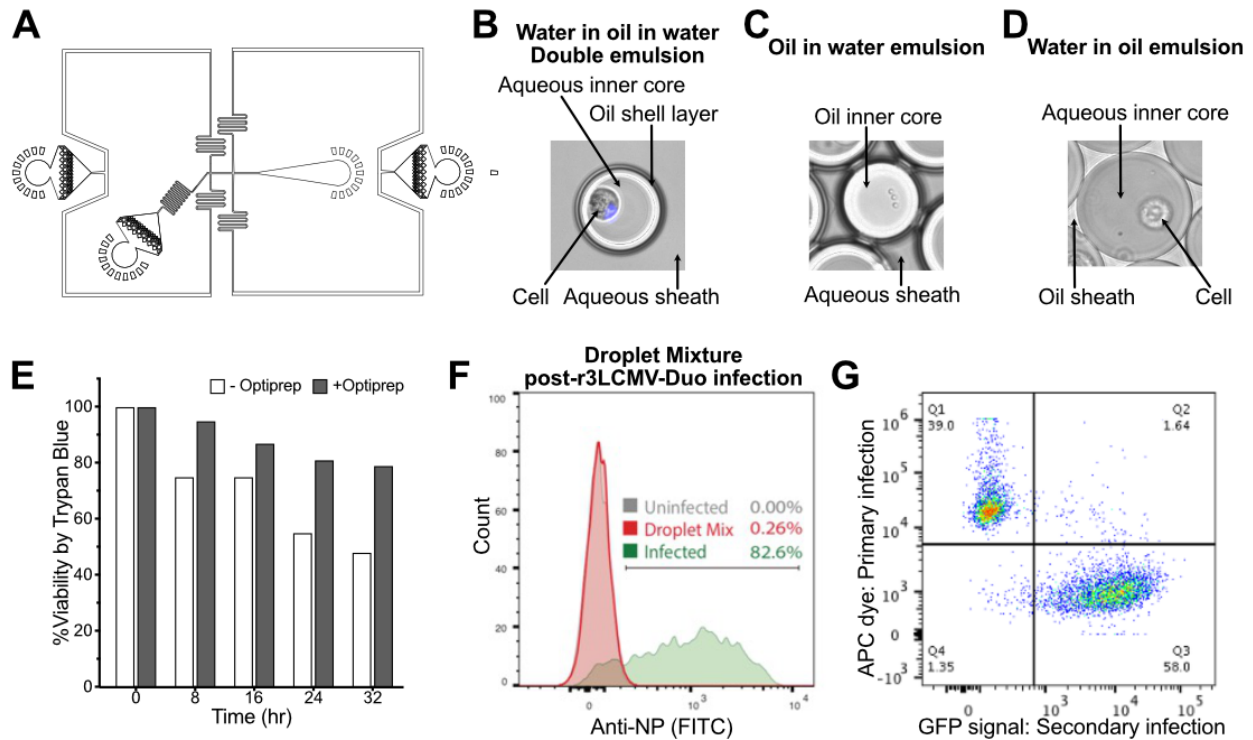


Figure 3.3 Droplet-based microfluidics allows for isolated viral infection events within one tube. (A) Schematic of a droplet-generating microfluidic device with two inlets. (B) Representative brightfield microscopy image of a double emulsion droplet with an aqueous sheath, oil shell layer, and an aqueous inner core around a cell. (C) Representative brightfield microscopy image of an oil-in-water single emulsion with an aqueous sheath and an oil inner core. (D) Representative brightfield microscopy image of a water-in-oil emulsion with an oil sheath and an aqueous inner core surrounding a cell. (E) Cell viability, as measured by trypan blue, over time in the presence or absence of cell density-matched amount of Optiprep. (F) Flow cytometry measurement of infection using anti-LCMV-NP (FITC) demonstrates that encapsulated uninfected cells lack an anti-LCMV-NP signal, as does encapsulated cells mixed with encapsulated virus 24 hpi. Conversely, encapsulated infected cells demonstrate high anti-LCMV-NP signal. (G) Primary infected cells stained with APC dye completely separate from secondary infected cells by flow cytometry following 16-hour co-encapsulation incubation.

matching reagent Optiprep (Millipore Sigma) in $80\mu\text{m}$ droplets, we expect to be able to overcome these limitations by encapsulating the secondary reporter cells at a distribution of $\lambda = 1$. This effectively renders the double Poisson distribution of co-encapsulation into a single Poisson distribution problem.

Finally, to ensure that encapsulated droplets never mix in unintended ways and that droplet-based methods are appropriate for functional screens, we devised two experiments. First, we tested whether virions could freely enter and exit droplets, thereby rendering this droplet-based method identical to a bulk assay. Droplets that contain virions alone and droplets that contain cells alone were mixed and incubated together. After 24 hours, droplets were broken and cells were fixed and stained for LCMV nucleoprotein (NP) (anti-NP 1-1.3 mouse monoclonal antibody, anti-mouse Alexa 488 secondary antibody). The cells recovered from the droplet mixture was nearly identical to the uninfected control in terms of LCMV NP staining, indicating that virions are trapped within the oil-water emulsion in which they are encapsulated (**Figure 3.3F**). Next, to verify that viral egress can occur within a droplet, infected primary cells, pre-stained with an CellTrace Far Red (Thermo Fisher Scientific), were co-encapsulated with uninfected secondary reporter cells at 8 hpi. Then, at 24 hpi, or 16 hours post-encapsulation, the cells were recovered from the droplet mixture, fixed, and stained for LCMV NP (**Figure 3.3G**). The resulting flow cytometry data clearly depicts two populations: an APC-positive and GFP-negative population (primary infection) and an APC-negative and GFP-positive population (secondary infection). Taken together, we believe that droplet-based microfluidics is a viable method to achieving high-throughput whole-genome functional screens in the context of viral infections, thereby allowing for elucidation of host factors that play a role during viral egress, a stage in the viral life cycle that has remained challenging to assay.

3.2.3 Novel droplet-based technologies allow for efficient sorting of double emulsions

The final step of a whole-genome functional screen is to select for a trait of interest such that gene perturbations of interest will achieve a higher level of signal than that of the background. For example, in a survival assay, perturbations of interest that are protective for the cells are selected and therefore represented in the final sequencing output [14,22]. In the case of viral protection, this selection would occur naturally when the virus being studied displays cytopathic effects in the cell line used [22,23]. Unfortunately, in the case of LCMV, and thus rLCMV-Duo, little cytopathic effect is observed in HeLaS3 cells. A second option to select for gene perturbations of interest would be to simply sort using a fluorescent reporter, which is what needs to be done here to verify successful (or unsuccessful) viral egress.

Here, the challenge would be attempting to sort fluorescent cells that are still encapsulated within their oil-water emulsion droplets. Traditional sorting methods such as fluorescence-activated cell sorting (FACS) will require re-encapsulating droplets into a double emulsion with an aqueous sheath fluid [24,25]. Standard FACS nozzles are often limited in sizes. Adding another layer to create a water in oil in water double emulsion (**Figure 3.3B**) may cause the diameter of the droplet to exceed that of what can fit through a standard FACS nozzle. Double emulsions have been successfully sorted on the Sony SH800 in 40 μ m diameter droplets, however HeLaS3 cells are unable to survive for 24 hours in such little media [24].

Another possibility for droplet sorting would be to explore microfluidic sorting options [26]. In this scenario, the cells could be sorted while in the state of a single emulsion, though the challenges would lie in the setup of the apparatus to be able to rapidly identify and separate cells depicting fluorescence of interest at a rate that is comparable to commercial FACS machines (over 10 000 events/second). Due to the artisanal nature of PDMS microfluidic cell sorters, achieving high levels of reproducibility while using microfluidic sorters remains an art.

Finally, should droplet sorting prove to be an insurmountable challenge, cross-linkage of primary and secondary infection cells prior to encapsulation would allow breakage of the emulsion prior to the sorting step. This can be achieved in a multitude of ways, including UV-activated linkers [27], chemical linkers [28], and even lipid-based oligonucleotide linkers [29]. By preemptively linking the primary and secondary infection cells together, the oil-water emulsion can be broken while still allowing cells originating from the same droplet to be sorted together.

3.3 IDENTIFICATION OF NOVEL HOST FACTORS INVOLVED IN VIRAL EGRESS

3.3.1 Droplet-based high-throughput whole-genome function screen for viral egress factors

To conduct a whole-genome functional screen to identify host factors for LCMV egress, we chose to use the humanV2 guide library generated by the Weismann lab at UCSF [30]. First, the single guide RNA (sgRNA) guide library is transduced into the HeLaS3 dC9K cell line (**Figure 3.4**). These cells, following a puromycin selection, would then be infected with rLCMV-Duo at a high multiplicity of infection (MOI). During this initial infection, the HeLaS3 dC9K or the primary infected cells, can be cultured in bulk for up to 8 hours, as viral egress has not yet occurred. Next, these primary infected cells will be co-encapsulated with the secondary infection reporter cell line, HeLaS3 dC9K GFP1-10, using a microfluidic droplet generator. While the primary infected cells are loaded at $\lambda = 0.1$ for single cells per droplet, the secondary reporter can be loaded at up to $\lambda = 3$ to ensure that every single drop produced would have at least one secondary reporter cell. These droplets are then incubated for an additional 16 hours prior to sorting (24 hpi, 16 hours post encapsulation).

Following viral egress into the secondary reporter cells, the cell pairs will then be sorted based on fluorescence. Droplets that contain both mCherry and GFP signals represents successful secondary infection, and thus do not represent gene perturbations that affect viral

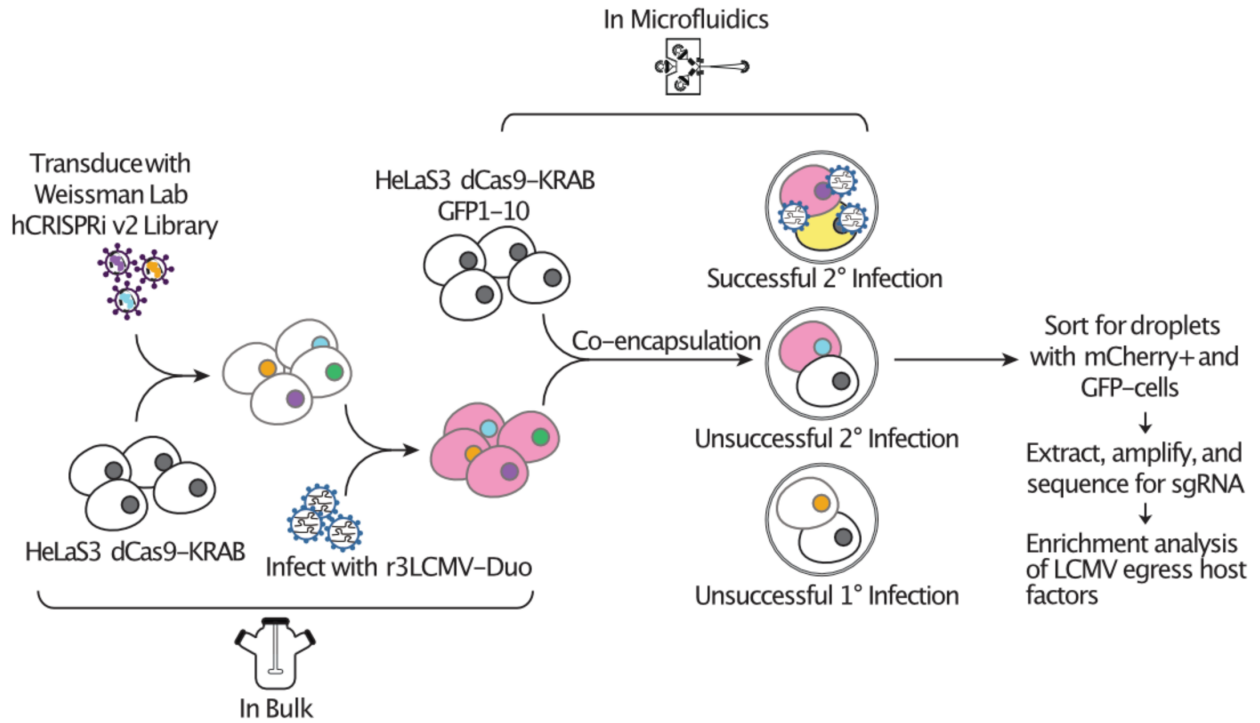


Figure 3.4 Genome-wide CRISPRi knockdown screen in human cells can be used to identify host factors important for LCMV budding and viral egress.

Schematic of CRISPR-based KD screen done in HeLaS3 cell line for identification for LCMV host factors using a droplet-based system.

egress, are discarded immediately. Droplets that contain neither mCherry nor GFP signals indicate an unsuccessful primary infection. Though this population is not one that this study is designed for, they can be sorted and sequenced to compare with the list of genes enriched from a standard bulk whole-genome screen [14,22]. Finally, droplets that contain only mCherry signal, but no GFP signal, should be sorted for genomic extraction. The sgRNAs present in this population represent gene perturbations that disrupted viral egress and should be PCR amplified and identified via next-generation sequencing. MAGeCK algorithm [31] can then be used to calculate a robust rank aggregation and produce a significance score, called the MAGeCK score, that describes the enrichment of genes important for viral egress above an unsorted background.

3.3.2 Candidate-based high-throughput function screen for viral egress factors

Should an appropriate method for sorting large droplets cannot be achieved, the designed reporter system for assaying successful viral egress can still be utilized in a candidate-based approach. Once again, HeLaS3 dC9K cells can be transduced with the humanV2 guide library prior to puromycin selection [30]. Then, following infection with rLCMV-Duo at a high MOI, these primary infected cells can be sorted using a standard commercial FACS machine into 96-, 384-, or 1536-well plates that already have secondary infection reporter cells seeded inside. Following a 24-hour incubation to allow time for virions to be produced and egress, each well can be assayed using live fluorescence microscopy in an environmental chamber. If a well contains cells with mCherry signal but no GFP signal, then these cells can be expanded and ultimately lysed for genomic sequencing. From this specific population, the sgRNA guide identities can be recovered through sequencing. For a candidate-based approach, it may not even be necessary to compare against a background diversity, as each expanded well of cell

pairs depicting unsuccessful viral egress can be considered a gene of interest enriched above background already. If these cells were re-pooled for bulk expansion, then perhaps further analysis using MAGeCK may be required.

3.3.3 Expected results

Ultimately, for cell pairs that that resulted in no mCherry signal at all, indicating that no primary infection occurred, we would expect the results of this population to look very similar to that of a bulk whole-genome CRISPRi functional genomics screen. Genes that are known to be play a role in viral entry, such as DAG1 and CD164, are expected to be enriched, as well as many members of the COG family, V-type ATPases, and heparin biosynthesis pathway [14,22,32]. We may also observe an enrichment of guides corresponding to eukaryotic transcription factors such as EIF4A1, which may have played a role in mCherry expression [33–37]. Control guides such as non-targeting guides and guides corresponding to genes such as GAPDH are not expected to experience enrichment.

In the viral egress screen, on the other hand, we expect a completely different set of genes to be enriched. These genes should correspond to findings from previous RNAi and candidate-based studies that focus on LCMV budding and assembly [35,38]. Host factors associated secretory pathways (TSG101) and post-translational modifications (Ski-1/S1P), both of which are necessary for viral assembly of cleaved, mature LCMV glycoprotein, are expected to be enriched in this screening modality [13,38,39]. Once again, control guides corresponding to genes such as GAPDH or targeting sequences absent from the human genome (non-targeting guides) are not expected to experience enrichment.

3.4 CURRENT CHALLENGES, IMPACT, AND FUTURE DIRECTIONS

Although we have made great strides in developing the technology necessary for conducting a whole-genome functional screen specifically targeting the viral egress life stage, there remains certain steps that require additional optimization. Current challenges include resolving our inability to sort double emulsion droplets at sizes that allow HeLaS3 cells to remain viable for over 24 hours. Although double emulsion sorting techniques exist [24,25], they require the inner aqueous core to be less than $40\mu\text{m}$ in diameter. Previously, we have demonstrated that HeLaS3 cells can remain viable for over 24 hours in $80\mu\text{m}$ droplets. However, the 2-fold decrease in diameter translates to an 8-fold decrease in media volume, leading to rapid cell death due to lack of nutrients and overcrowding.

An alternative path to explore would be to work with smaller suspension cell types than HeLaS3 such that there is a greater likelihood of surviving encapsulation overnight [40]. Another route to explore would be to entertain the idea of using a different virus, specifically one that is able to egress following a much shorter incubation period [41]. All in all, there are many other pathways to explore to develop this idea into a mature technology that can be used to study viral egress in a host of different cell types with a variety of different viruses.

While current technological developments have allowed us to identify novel host determinants of viral infections in an unbiased, high-throughput manner, one aspect of the viral life cycle remains elusive. Viral egress, the stage in which mature virions are assembled and bud off the infected host cells, remain challenging to study in isolation. By developing a high-throughput functional genomic screening assay that specifically targets exploration of the viral egress stage, we hope to address some of the lack of understanding of fundamental host factors that impact LCMV infection. Knowledge gained through the characterization of host determinants for viral egress can be used inform future therapies as well.

3.5 REFERENCES

1. Han J, Perez JT, Chen C, et al. Genome-wide CRISPR/Cas9 Screen Identifies Host Factors Essential for Influenza Virus Replication. *Cell Rep* **2018**; 23:596–607.
2. Marceau CD, Puschnik AS, Majzoub K, et al. Genetic dissection of Flaviviridae host factors through genome-scale CRISPR screens. *Nature* **2016**; 535:159–163.
3. Park RJ, Wang T, Koundakjian D, et al. A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors. *Nat Genet* **2017**; 49:193–203.
4. Savidis G, McDougall WM, Meraner P, et al. Identification of Zika Virus and Dengue Virus Dependency Factors using Functional Genomics. *Cell Rep* **2016**; 16:232–246.
5. Zhang R, Miner JJ, Gorman MJ, et al. A CRISPR screen defines a signal peptide processing pathway required by flaviviruses. *Nature* **2016**; 535:164–168.
6. Staller E, Sheppard CM, Neasham PJ, et al. ANP32 Proteins Are Essential for Influenza Virus Replication in Human Cells. *J Virol* **2019**; 93:e00217-19.
7. Hosmillo M, Lu J, McAllaster MR, et al. Noroviruses subvert the core stress granule component G3BP1 to promote viral VPg-dependent translation. *Elife* **2019**; 8:e46681.
8. Knipe DM, Howley PM, editors. *Fields virology*. 6th ed. Philadelphia, PA: Wolters Kluwer/Lippincott Williams & Wilkins Health, 2013.
9. Lechartier B, Rybniker J, Zumla A, Cole ST. Tuberculosis drug discovery in the post-post-genomic era. *EMBO Mol Med* **2014**; 6:158–168.
10. Ford K, McDonald D, Mali P. Functional Genomics via CRISPR-Cas. *J Mol Biol* **2019**; 431:48–65.
11. Jackrel ME, Cortajarena AL, Liu TY, Regan L. Screening libraries to identify proteins with desired binding activities using a split-GFP reassembly assay. *ACS Chem Biol* **2010**; 5:553–562.

12. Emonet SF, Garidou L, McGavern DB, de la Torre JC. Generation of recombinant lymphocytic choriomeningitis viruses with trisegmented genomes stably expressing two additional genes of interest. *Proc Natl Acad Sci U S A* **2009**; 106:3473–3478.
13. Kunz S, Edelmann KH, de la Torre J-C, Gorney R, Oldstone MBA. Mechanisms for lymphocytic choriomeningitis virus glycoprotein cleavage, transport, and incorporation into virions. *Virology* **2003**; 314:168–178.
14. 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* **2022**; :e0020522.
15. Collins DJ, Neild A, deMello A, Liu A-Q, Ai Y. The Poisson distribution and beyond: methods for microfluidic droplet production and single cell encapsulation. *Lab Chip* **2015**; 15:3439–3459.
16. Mazutis L, Gilbert J, Ung WL, Weitz DA, Griffiths AD, Heyman JA. Single-cell analysis and sorting using droplet-based microfluidics. *Nat Protoc* **2013**; 8:870–891.
17. Clausell-Tormos J, Lieber D, Baret J-C, et al. Droplet-based microfluidic platforms for the encapsulation and screening of Mammalian cells and multicellular organisms. *Chem Biol* **2008**; 15:427–437.
18. Shembekar N, Hu H, Eustace D, Merten CA. Single-Cell Droplet Microfluidic Screening for Antibodies Specifically Binding to Target Cells. *Cell Rep* **2018**; 22:2206–2215.
19. Klein AM, Mazutis L, Akartuna I, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **2015**; 161:1187–1201.
20. Melin J, Quake SR. Microfluidic Large-Scale Integration: The Evolution of Design Rules for Biological Automation. *Annu Rev Biophys Biomol Struct* **2007**; 36:213–231.
21. Ward T, Faivre M, Abkarian M, Stone HA. Microfluidic flow focusing: drop size and scaling in pressure versus flow-rate-driven pumping. *Electrophoresis* **2005**; 26:3716–3724.

22. Bakkers MJG, Moon-Walker A, Herlo R, et al. CD164 is a host factor for lymphocytic choriomeningitis virus entry. *Proc Natl Acad Sci U S A* **2022**; 119:e2119676119.
23. 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.
24. Brower KK, Khariton M, Suzuki PH, et al. Double Emulsion Picoreactors for High-Throughput Single-Cell Encapsulation and Phenotyping via FACS. *Anal Chem* **2020**; 92:13262–13270.
25. Zinchenko A, Devenish SRA, Kintsjes B, Colin P-Y, Fischlechner M, Hollfelder F. One in a million: flow cytometric sorting of single cell-lysate assays in monodisperse picolitre double emulsion droplets for directed evolution. *Anal Chem* **2014**; 86:2526–2533.
26. Agresti JJ, Antipov E, Abate AR, et al. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc Natl Acad Sci U S A* **2010**; 107:4004–4009.
27. Berg M, Michalowski A, Palzer S, Rupp S, Sohn K. An in vivo photo-cross-linking approach reveals a homodimerization domain of Aha1 in *S. cerevisiae*. *PLoS One* **2014**; 9:e89436.
28. Kim E, Koo H. Biomedical applications of copper-free click chemistry: in vitro, in vivo, and ex vivo. *Chem Sci* **2019**; 10:7835–7851.
29. 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.
30. Horlbeck MA, Gilbert LA, Villalta JE, et al. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife* **2016**; 5:e19760.
31. 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.
32. Volland A, Lohmüller M, Heilmann E, Kimpel J, Herzog S, von Laer D. Heparan sulfate proteoglycans serve as alternative receptors for low affinity LCMV variants. *PLoS Pathog* **2021**; 17:e1009996.

33. Jae LT, Raaben M, Riemersma M, et al. Deciphering the glycosylome of dystroglycanopathies using haploid screens for lassa virus entry. *Science* **2013**; 340:479–483.
34. Raaben M, Jae LT, Herbert AS, et al. NRP2 and CD63 Are Host Factors for Lujo Virus Cell Entry. *Cell Host Microbe* **2017**; 22:688-696.e5.
35. Pasqual G, Rojek JM, Masin M, Chatton J-Y, Kunz S. Old world arenaviruses enter the host cell via the multivesicular body and depend on the endosomal sorting complex required for transport. *PLoS Pathog* **2011**; 7:e1002232.
36. Hulseberg CE, Fénéant L, Szymańska KM, White JM. Lamp1 Increases the Efficiency of Lassa Virus Infection by Promoting Fusion in Less Acidic Endosomal Compartments. *mBio* **2018**; 9:e01818-17.
37. King BR, Hershkowitz D, Eisenhauer PL, et al. A Map of the Arenavirus Nucleoprotein-Host Protein Interactome Reveals that Junín Virus Selectively Impairs the Antiviral Activity of Double-Stranded RNA-Activated Protein Kinase (PKR). *J Virol* **2017**; 91:e00763-17.
38. Perez M, Craven RC, de la Torre JC. The small RING finger protein Z drives arenavirus budding: implications for antiviral strategies. *Proc Natl Acad Sci U S A* **2003**; 100:12978–12983.
39. Beyer WR, Pöpplau D, Garten W, von Laer D, Lenz O. Endoproteolytic processing of the lymphocytic choriomeningitis virus glycoprotein by the subtilase SKI-1/S1P. *J Virol* **2003**; 77:2866–2872.
40. El Debs B, Utharala R, Balyasnikova IV, Griffiths AD, Merten CA. Functional single-cell hybridoma screening using droplet-based microfluidics. *Proc Natl Acad Sci U S A* **2012**; 109:11570–11575.
41. Passalacqua KD, Lu J, Goodfellow I, et al. Glycolysis Is an Intrinsic Factor for Optimal Replication of a Norovirus. *mBio* **2019**; 10:e02175-18.

4 PERFORMANCE CHARACTERISTICS OF A RAPID SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS 2 ANTIGEN DETECTION ASSAY AT A PUBLIC PLAZA TESTING SITE IN SAN FRANCISCO

4.1 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).

4.2 INTRODUCTION

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 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.

4.3 METHODS

4.3.1 *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.

4.3.2 *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].

4.3.3 *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.

4.3.4 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 μ L 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.

4.3.5 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 μ L of N protein was absorbed onto the provided swab sample.

4.3.6 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.

4.4 RESULTS

4.4.1 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 4.1**). 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 Ct values (average of N and E genes) of 30.3 and 28.8, respectively, in this assay.

4.4.2 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.

4.4.3 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 4.2**). 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 4.2A**). 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 4.2B**). In each case, the corresponding image is shown to demonstrate the correspondence between RT-PCR and the visual result (**Figure 4.2C**).

4.4.4 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 1.9×10^4 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 (2.6×10^3 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 ($>10^4$), 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.

4.4.5 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 4.S1**). 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 4.S1A**).

4.5 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.

4.6 ACKNOWLEDGEMENTS

We thank Bevan Dufty and the BART team, Jeff Tumlin and the San Francisco MUNI, Supervisor Hillary Ronen, Mayor London Breed, Grant Colfax, MD and the Department of Public Health, Salu Ribeiro, MS and Bay Area Phlebotomy and Laboratory services, the PrimaryBio testing platform for coronavirus disease 2019, and our community ambassadors and volunteers. We also thank Don Ganem, MD for writing assistance and critical discussion, Andreas Puschnik, PhD for Huh7.5.1 overexpression cells used for severe acute respiratory syndrome coronavirus disease 2 growth, and Terry Robins, PhD, Stephen Kovacs, and John Hackett Jr, PhD from Abbott Laboratories for their support. Finally, we thank both Abbott Laboratories and

the California Department of Public Health for their generous donations of BinaxNOW COVID-19 Ag cards.

This study was supported by the University of California, San Francisco, the Chan Zuckerberg Biohub, the Chan Zuckerberg Initiative, the San Francisco Latino Task Force, the National Institute of Allergy and Infectious Diseases (grants T32 AI060530 to L. R. and F31AI150007 to S. S.), and a private donor.

No authors reported conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

4.7 FIGURES

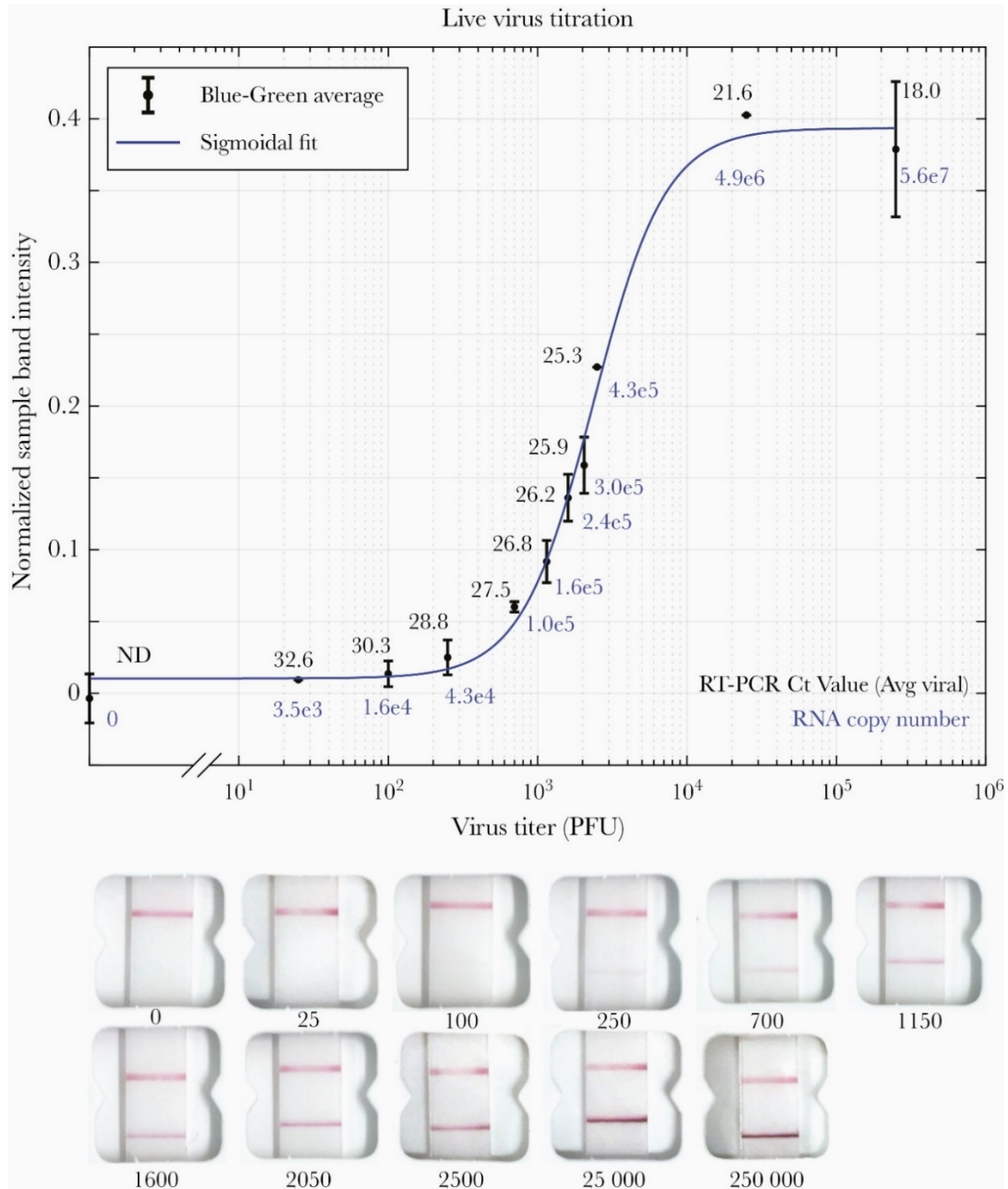


Figure 4.1 Titration of in vitro grown severe acute respiratory syndrome coronavirus 2 and detection with Binax-CoV2 assay.

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.

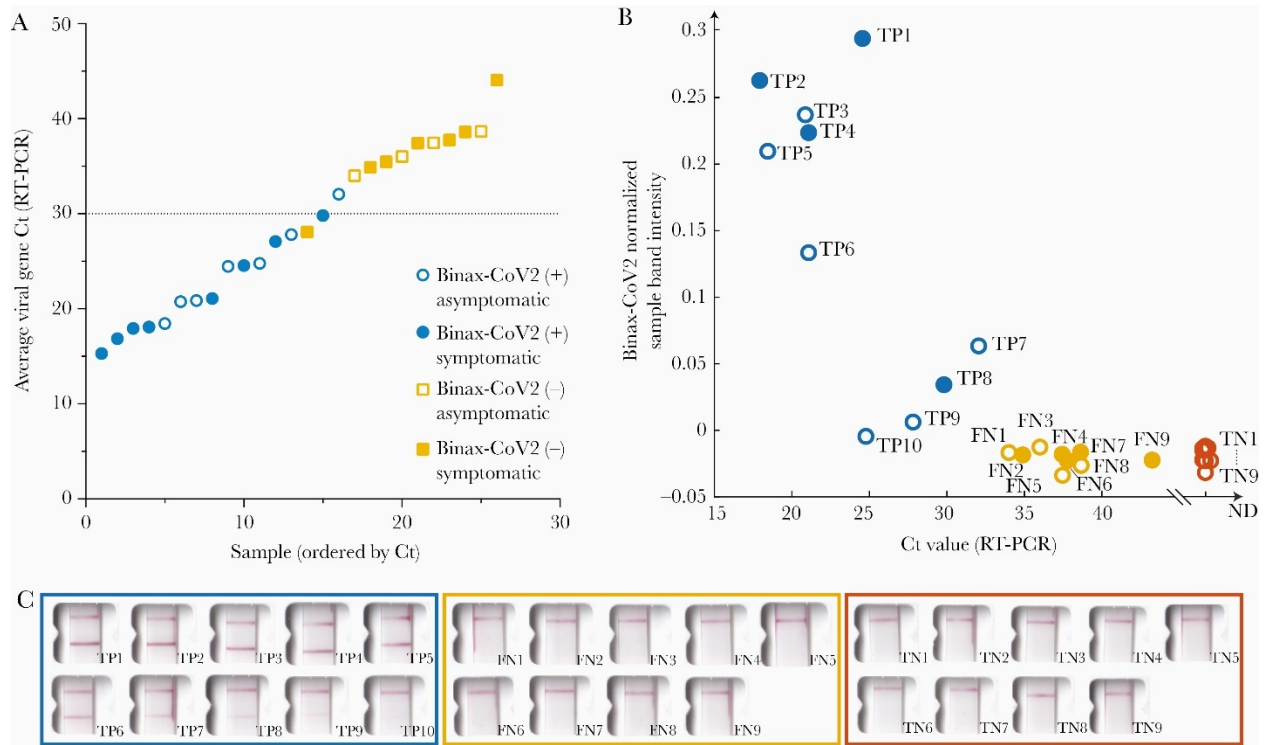


Figure 4.2 Comparison of Binax-CoV2 test with quantitative reverse-transcription polymerase chain reaction (RT-PCR) test

(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).

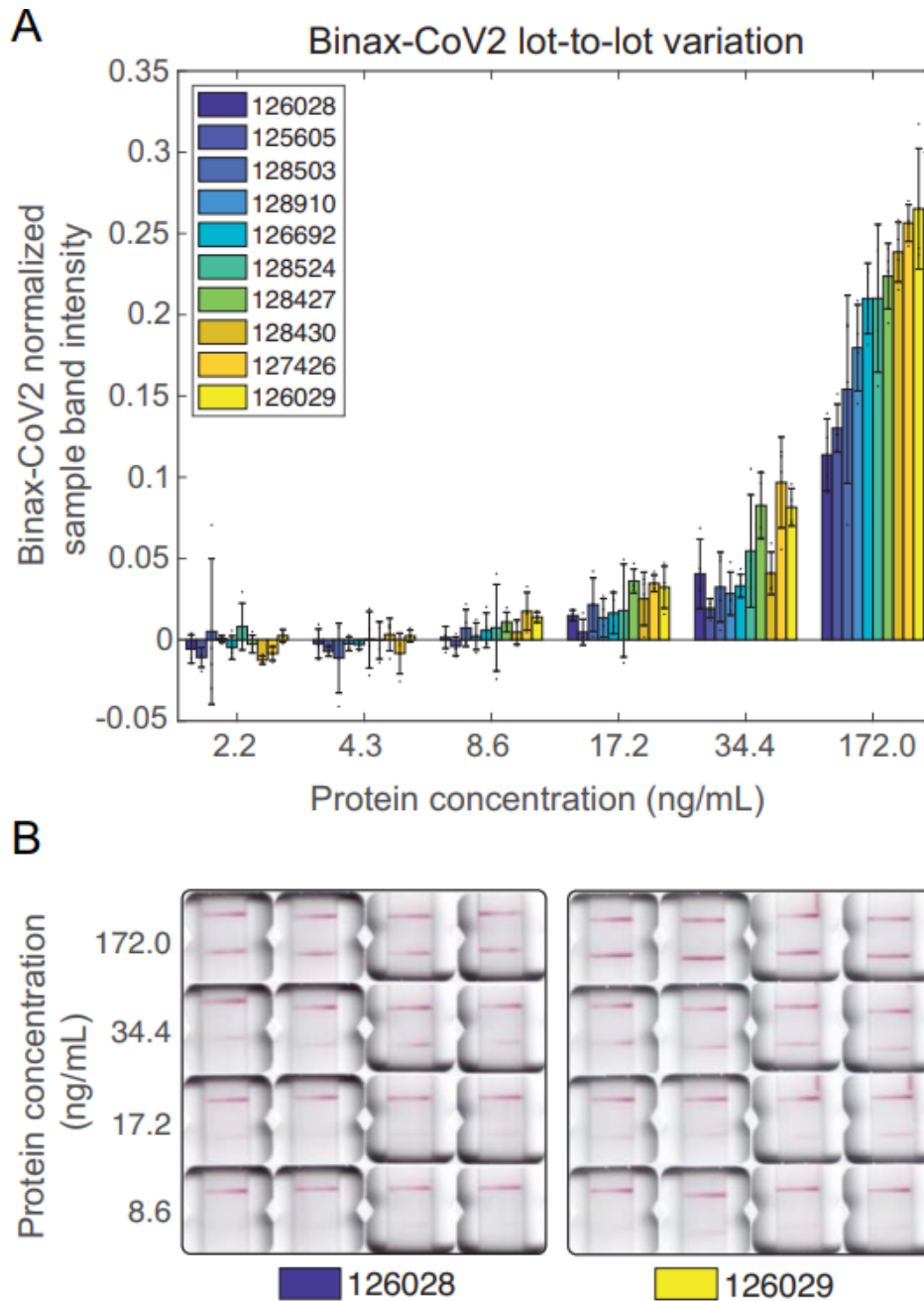


Figure 4.S1 Variability of signal intensity in Binax-CoV2 card lots

(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.

4.8 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://corona-virus.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/manu-script/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.

5 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

5.1 ABSTRACT

Background. Sequencing of the 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 November 22-December 1, 2020 and January 10-29, 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 1,099 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, RR=1.28; 95%

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.

Summary: We observed a growing prevalence and modestly elevated attack rate for “West Coast” SARS-CoV-2 variants in a community testing setting in San Francisco during January 2021, suggesting its modestly higher transmissibility.

5.2 INTRODUCTION

Genomic surveillance during the 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 (e.g., COVID-19 Genomics UK Consortium). Examples include the rapidly spreading B.1.1.7 lineage documented in the UK and the B.1.351 lineage described from South Africa, or the P.1/P.2 lineages that harbor the spike E484K mutation which is associated with reduced neutralization in laboratory experiments [2–5].

In the US, genomic surveillance is sparse relative to the number of confirmed cases (27.8 million as of Feb 20, 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 highlights the utility of this approach in the US [9].

Genomic sequencing of SARS-CoV-2 in California has predominantly been conducted by academic researchers and non-profit biomedical research institutions (for example, 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. While 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 (2,309/4,305; 53.6%; GISAID accessed February 20, 2021).

The majority of sequencing efforts in the US utilize samples from symptomatic individuals or outbreaks, introducing selection bias making interpretation of trends, such as the rise in lineage prevalence, complex. Further, 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 COVID-19 incidence, during two periods: November 22-December 1, 2020 and January 10-29, 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.

5.3 METHODS

5.3.1 Study setting and population

Over November 22 to December 1, 2020 and January 10-29, 2021, BinaxNOW™ rapid antigen tests (BinaxNOW) 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, 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].

5.3.2 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 can be found in 4.7 Supplementary Methods.

5.3.3 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% 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].

5.3.4 Bayesian Phylogenetic Analysis

We compared the growth rates of B.1.427 and B.1.429 PANGO lineages against two 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 can be found in 4.7 Supplementary Methods.

5.3.5 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.

5.4 RESULTS

5.4.1 Low-Barrier SARS-CoV-2 Testing and Sequencing

From November 22 to December 1, 2020, 3,302 rapid direct antigen tests were performed on 3,122 unique individuals; sample characteristics from this testing have been previously described [11]. From January 10-29, using identical methods, 8,822 rapid direct antigen tests were performed on 7,696 unique individuals, representing 5,239 households; household attack rate analyses were restricted to January samples, described here (**Table 5.S1**).

Test subjects originated from addresses in 8 Bay Area counties, indicating a wide catchment area (**Figure 5.1**). 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% (**Table 5.S2**, sequences deposited in GISAID). These 986 samples, together with an additional 191 SARSCoV-2 genome sequences generated from the same testing site during the period of November 22-December 1, 2020 [11,19] had adequate coverage of the full genome or spike protein for further

analysis based on S gene sequence (**Table 5.S3**). 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 (**Figure 5.S1**). 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 re-infection 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 5.2**). These lineages are separated by differing mutations ORF1a and ORF1b, including ORF1b:P976L and ORF1a:I4205V, respectively. Sequencing of 191 viral genomes from November 22-December 1, 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 5.2A-B**), 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 5.2C-D**). This increase in frequency is consistent with an expansion of viruses more broadly in California carrying these same mutations [21].

Additional non-synonymous 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 5.2E, Table 5.S3**). 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 [22–25]. 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 two 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].

5.4.2 Disease Severity

The SARS-CoV-2 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; IQR 6.4) versus non-West Coast (mean Ct 23.67; IQR 7.8) strains (95% CI: -0.77-0.50, p-value = 0.67) (Figure 4.S2, Table 4.S2). 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).

5.4.3 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. (**Table 5.1, Table 5.S4**).

The 319 index cases had a total of 1,241 non-index 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 B.1.427; 40.3%, 79/196 for B.1.429), while 29.4% (122/415) of contacts exposed to non-West Coast variant tested positive (**Table 5.2**). 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 = 0.05). 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: 0.89-1.59, p-value = 0.20) and 1.43 (95% CI: 1.07-1.91, p-value = 0.02), respectively. Dispersion ratios were greater than 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 (aRR: 1.25, 95% CI: 0.98-1.59, p-value: 0.07 for West Coast vs non-West Coast variants; aRR: 1.19, 95%CI: 0.90-1.59, p-value = 0.20 and aRR: 1.36, 95%CI: 1.01-1.83, p-value = 0.04 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 5.3**); 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 (**Table 5.S5**).

5.4.4 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 two 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: 0.98-1.05) 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.

5.5 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 [21].

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 comprise 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. While the frequency of both lineages increased in this study and in California more widely [21], 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 one. While 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 [26], 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 (**Figure 5.S2**), 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 co-morbidities 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 Feb 24, 2021), suggesting that introductions of B.1.1.7 have been rare

in this catchment area, despite high SARS-CoV-2 incidence [27]. 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; further, 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 under- (or over-) estimate of secondary attack rate, and while 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. While 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.

5.6 ACKNOWLEDGEMENTS

The authors would like to thank the hundreds of academic labs and public health institutions that have been sequencing and publicly depositing genomic sequences, analysis tools, and epidemiological data throughout the pandemic. They thank Bevan Dufty and the BART team, Jeff Tumlin and the San Francisco MUNI, Supervisor Hillary Ronen, Mayor London Breed, Dr Grant Colfax and the Department of Public Health, Salu Ribeiro and Bay Area Phlebotomy and Laboratory services, PrimaryBio COVID testing platform, and our community ambassadors and volunteers. They would like to thank the Chan Zuckerberg Initiative, Greg and Lisa Wendt, Anne-Marie and Wylie Peterson, Gwendolyn Holcombe, Andrew and Dokleida Kawaja, Michael and Hazel Kawaja, Chris and Mitchell Kawaja, Carl Kawaja, the McKinnon Family Foundation, and Kevin and Julia Hartz, for their critical support and input. We also thank Jack Kamm, Peter Kim, Don Ganem, Sandy Schmidt, Cori Bargmann, Norma Neff, and Christopher Hoover for technical assistance and discussion.

This work was supported by the UCSF COVID Fund (to J. D., D. H., J. L., and M. L.), the National Institutes of Health (grant numbers UM1AI069496 to J. D. and F31AI150007 to S. S.), the Chan Zuckerberg Biohub (to J. D. and D. H.), the Chan Zuckerberg Initiative (to J. D. and D. H.), and a group of private donors. The BinaxNOW cards were provided by the California Department of Public Health.

J. D. is a member of the scientific advisory board of The Public Health Company, Inc., and is a scientific advisor for Allen & Co. J. D. also reports options granted for service on the Scientific Advisory Board of The Public Health Company. D. H. reports non-financial support from Abbott, outside the submitted work. P. A. and D. H. report funding from Chan Zuckerberg Biohub/Initiative. P. A. and A. M. are employed by Chan Zuckerberg Biohub. C. M. reports grant funding from the NIH and gift funds from the Stupski foundation. All other authors report no potential conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

5.7 SUPPLEMENTARY METHODS

5.7.1 SARS-CoV-2 genomic sequence recovery

Swab samples from individuals testing positive by BinaxNOW were placed in DNA/RNA shield and processed as previously described [30]. Extracted total nucleic acid was diluted based on average SARS-CoV-2 N and E gene cycle threshold (Ct) values; samples with a Ct range 12-15 were diluted 1:100, 15-18 1:10 and >18 no dilution. For high throughput scaling, library preparation reaction volumes and dilutions were miniaturized utilizing acoustic liquid handling. Library preparation followed either the modified versions of the Primal-Seq Nextera XT version 2.0 protocol [31,32], or the modified version of SARS-CoV-2 Tailed Amplicon Illumina Sequencing V.2 [33], both using the ARTIC Network V3 primers [34]. A subset of initial samples were library prepared using the Tailed Amplicon Sequencing V.2 with only primer pairs 71-84 of the ARTIC V3 primers to tile all of the S gene. Final libraries were sequenced by paired-end 2 x 150bp sequencing on an Illumina NovaSeq platform, or for the S-gene-only set, 2 x 300bp on an Illumina MiSeq.

5.7.2 SARS-CoV-2 consensus genome generation

Raw FASTQ files were imported into IDseq and consensus genomes were generated automatically using the embedded SARS-CoV-2 pipeline [16]. Specifically, minimap2 was used to align raw reads to the reference genome MN908947.2 [35], then the consensus sequence was generated using samtools [36], mpileup and ivar [37]. The IDseq consensus genome

pipeline is implemented in WDL [38]. Viral genomes with at least 92% (27,500nt) recovery were uploaded to GISAID [19], a worldwide repository for SARS-CoV-2 genomes and Genbank [39]. Phylogenetic analysis was conducted and results were visualized in Nextclade (<https://clades.nextstrain.org>) [21].

5.7.3 Bayesian Phylogenetic Analysis

To compare the viral diversity of the two variants of interest, B.1.427 and B.1.429, we identified two other SARS-CoV-2 lineages, B.1.232 and B.1.243 that were prevalent in the state of California from July 2020 onwards. Both of these lineages contained more sequences on GISAID from California than any other location, based on the PANGO lineage assignment [40] on GISAID. For each of the 4 lineages spreading in California, we randomly subsampled all available genomes from GISAID and this study to 500 or fewer genomes. For subsampled genomes, we aligned them against the reference genome (Genbank accession: MN996528.1) using MAFFT v7.471 [41] with default settings. Each multi-sequence alignment was used to build a separate maximum likelihood tree in IQ-TREE v.1.6.12 [42] with default options. The trees were rooted at the reference genome. The maximum likelihood tree was used to visually identify outlier sequences which could have been misclassified as that PANGO lineage. The resulting number of genomes included in the downstream analysis were B.1.232: 368; B.1.243: 500; B.1.427: 495; B.1.429: 443. The multi-sequence alignment of the coding region for each lineage was analyzed in BEAST v.1.10.4 [43] with unlinked molecular clocks between the S gene and other genes, uncorrelated relaxed molecular clock (lognormal distribution), GTR substitution model with 4 rate categories (selected by the BIC value in ModelTest [44]), and the Bayesian Skygrid population model [45]. Default prior values and operator values were used. The MCMC chains were 100M in length, sampled every 50,000, and the first 50% of samples

discarded as burnin. We ran 2 replicate MCMC chains for each analysis and used all samples to summarize the results.

We fit an exponential model to the median SkyGrid estimate between 2020-07-01 and 2021-01-01 to calculate the growth rate per day r , and calculated the reproductive number R using the formula [46] $R = (1+r/b)^a$, where $a=1.39$ is the shape parameter and $b=0.14$ is the scale parameter of a gamma generation time distribution with a mean of 5 days a standard deviation of 1.9 days [47]. The 95% confidence interval (95% CI) around R was calculated based on the 95% CI around the r estimate. Samples used for the Bayesian Analysis are detailed in the supplementary file.

5.8 FIGURES

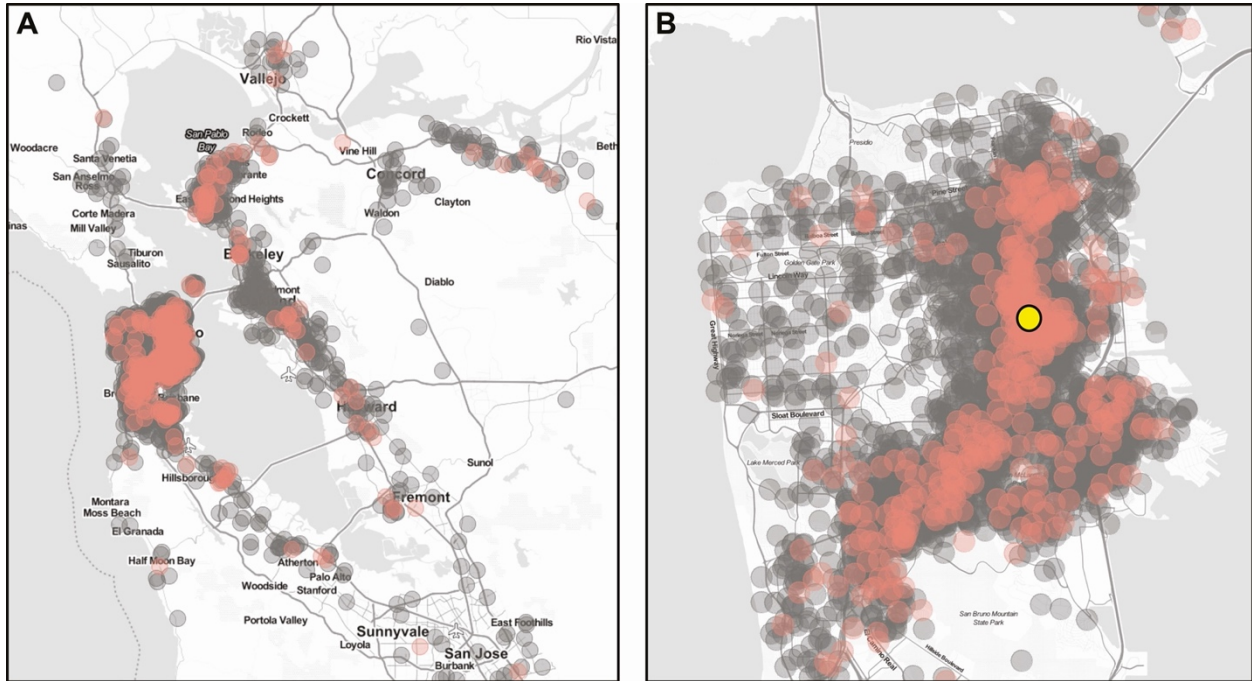


Figure 5.1 Testing catchment area

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 resides within San Francisco County (B), Map tiles by Stamen Design and data by OpenStreetMap.

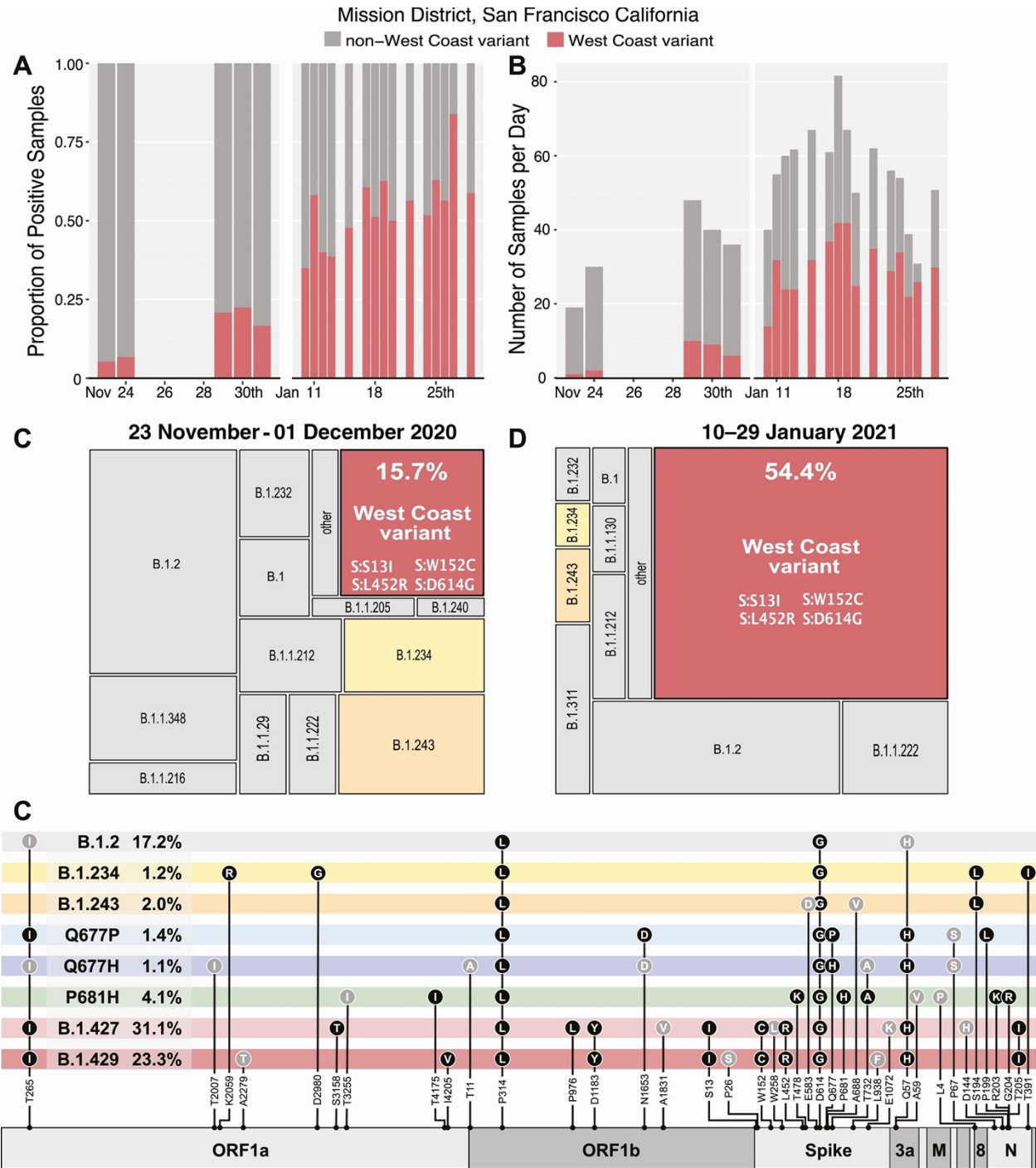


Figure 5.2 Variants observed at 24th & Mission

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.

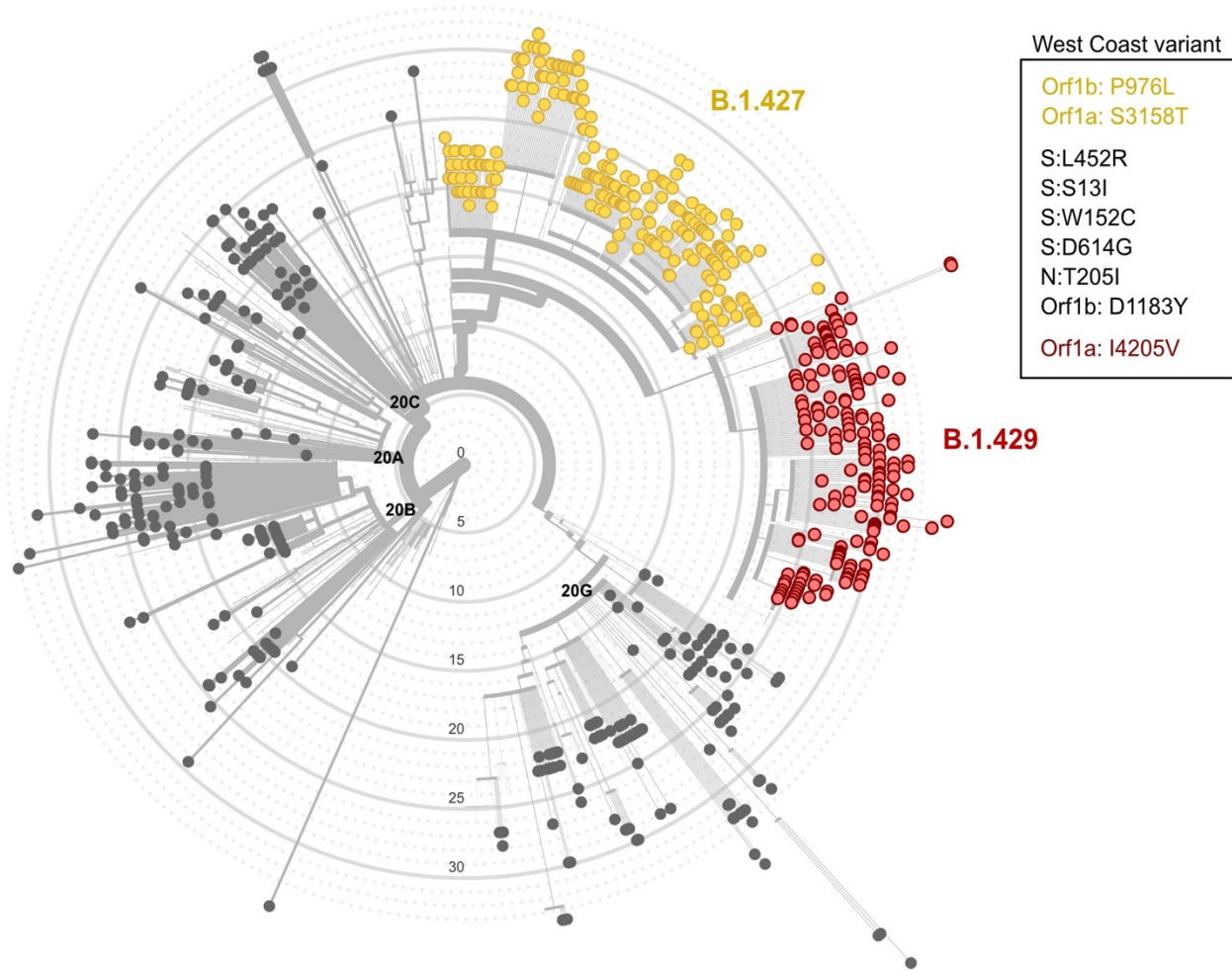


Figure 5.S1 NextClade Phylogeny

Radial tree phylogeny as visualized via NextClade [21] using full length genomes (n=614) from the Jan 10th - 29th sampling period and the San Francisco public reference tree (downloaded Feb 17th, 2021). For the purposes of visualization, six genomes were removed due to excessive private mutations. Colored nodes correspond to specific mutations as indicated in the figure legend.

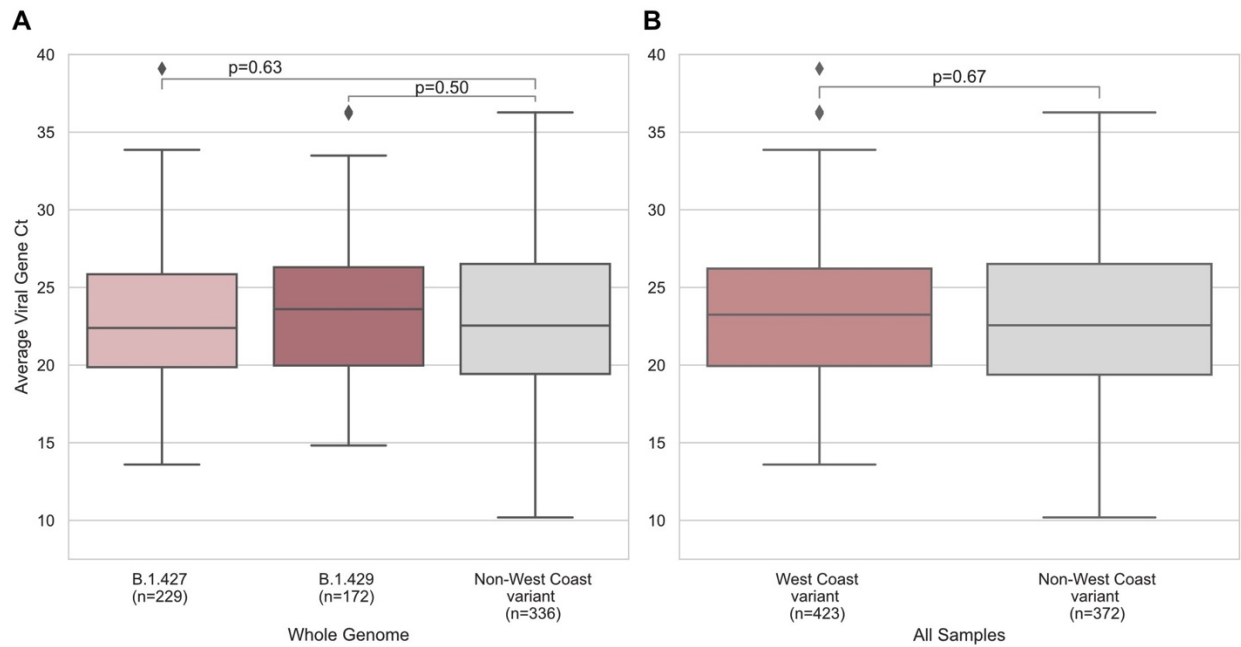


Figure 5.S2 Viral load by Ct value

No significant differences in mean viral load as measured by RT-qPCR Cts was measured between West Coast and non-West Coast variants in **(A)** the subset of samples that had only spike gene sequenced (95% CI: -1.97-1.29, p-value = 0.68), **(B)** the subset of samples that had undergone whole genome sequencing thus allowing for determination of West Coast variant PANGO lineage of either B.1.427 (95% CI: -1.84-0.25, p-value = 0.18) or B.1.429 (95% CI: -0.49-1.76, p-value = 0.38), **(C)** all samples sequenced in this study (95% CI: -0.63-0.83, p-value = 0.79).

Table 5.1 Characteristics of Households Included in the Household Attack Rate Analysis, Stratified by Strain

Race/Ethnicity (most common in household)	Non-West Coast		West coast		Total (N=319)
	(N=156)	B.1.427 (N=90)	B.1.429 (N=65)	All West coast (N=163) ^a	
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%)	2 (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%)	232 (73.0%)
Outside San Francisco	38 (24.4%)	19 (21.1%)	26 (40.0%)	48 (29.4%)	86 (27.0%)
Household Size					
2 persons	13 (8.3%)	12 (13.3%)	5 (7.7%)	20 (12.3%)	34 (10.7%)
3-4 persons	64 (41.0%)	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]

^a8 households with S gene only sequence available.^bHousehold density missing for 17 household

Table 5.2 Secondary Household Attack Rates for West Coast Variants, Combined and Disaggregated by B.1.427 and B.1.429

Class	Positives Among Tested Contacts (%)	Mean Household Attack Rate (95% CI)	Unadjusted		Adjusted	
			RR (95%CI)	p-value	aRR _{robust}	p-value
Non-West Coast	122/415 (29.4%)	25.6% (20.3-31)	-	-	-	-
West Coast	161/451 (35.7%)	35.9% (30.1-41.9)	1.28 (1.00-1.64)	0.05	1.25 (0.98-1.59)	0.07
Lineage						
B.1.427	78/235 (33.2%)	32.9% (25.4-40.6)	1.19 (0.89-1.59)	0.20	1.19 (0.90-1.59)	0.20
B.1.429	79/196 (40.3%)	40.9% (31.5-50.5)	1.43 (1.07-1.91)	0.02	1.36 (1.01-1.83)	0.04

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_{robust}, adjusted risk ratio; CI, confidence interval; RR, risk ratio

Table 5.3 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)	35/101 (34.7%)	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

Table 5.S1 Characteristics of all persons (N=8,822; 10.0% with positive result) tested at the community-based testing site between January 10-29

	Negative (N=7937)	Positive (N=885)	Total (N=8822)
Gender			
Female	3595 (90.2%)	391 (9.8%)	3986 (100%)
Male	4190 (89.7%)	481 (10.3%)	4671 (100%)
Other	152 (92.1%)	13 (7.9%)	165 (100%)
Age Group			
Age ≤ 12	488 (83.1%)	99 (16.9%)	587 (100%)
Age 13-35	3032 (88.8%)	382 (11.2%)	3414 (100%)
Age 36-64	3844 (91.5%)	355 (8.5%)	4199 (100%)
Age ≥ 65	573 (92.1%)	49 (7.9%)	622 (100%)
Race/Ethnicity			
Hispanic/Latinx	5700 (88.5%)	744 (11.5%)	6444 (100%)
Asian	610 (93.7%)	41 (6.3%)	651 (100%)
White/Caucasian	818 (96.3%)	31 (3.7%)	849 (100%)
Black or African American	228 (96.6%)	8 (3.4%)	236 (100%)
Other	581 (90.5%)	61 (9.5%)	642 (100%)
Occupation*			
Food & Beverage	987 (90.5%)	104 (9.5%)	1091 (100%)
Tradesperson	423 (87%)	63 (13%)	486 (100%)
Day laborer	227 (89.4%)	27 (10.6%)	254 (100%)
Healthcare	277 (93.9%)	18 (6.1%)	295 (100%)
Student	780 (84.4%)	144 (15.6%)	924 (100%)
Other	2557 (92.4%)	211 (7.6%)	2768 (100%)
No employment	1126 (88.8%)	142 (11.2%)	1268 (100%)
Day of Test Symptoms			
Asymptomatic	6089 (94.6%)	347 (5.4%)	6436 (100%)
Symptomatic	1848 (77.5%)	538 (22.5%)	2386 (100%)

* Occupation information missing from 1,736 persons.

Table 5.S2 Sample sequencing summary and viral cycle threshold characteristics

		Spike Gene Only	Whole Genome		Total Count (Jan 10-Jan 29)
Total Samples Run		80	775		855
Total High Quality (>92% coverage)		58	737		795
Average Reads/Sample		91,812	2,457,612		-
Average % Gene/Genome Covered		86.06%	97.48%		96.41%
Mean Viral Ct		24.46	23.53		23.62
Mean Viral Ct West Coast variant	B.1.427	25.46	23.44	22.97	23.56
	B.1.429			24.04	
Mean Viral Ct Non-West Coast variant		23.21	23.63		23.67
Median Viral Ct		24.53	23.14		23.26
Median Viral Ct West Coast variant	B.1.427	25.05	23.26	22.54	23.49
	B.1.429			23.98	
Median Viral Ct Non-West Coast variant		23.68	22.85		22.89
Viral Ct Interquartile Range (IQR)		8.43	7.09		7.25
Viral Ct IQR West Coast variant	B.1.427	6.51	6.43	6.28	6.42
	B.1.429			6.89	
Viral Ct IQR Non-West Coast variant		7.97	7.91		7.76

Table 5.S3 Amino acid mutations observed in the spike gene and count of sequences per mutation in each study

Spike domain	Position in S gene	Mutation	Count Nov. 23- Dec. 1 2020 (n=191)	Count Jan. 10- Jan.29 2021 (n=795)	Total count (n=986)	Covariant group key		
	5	L5F	4	5	9			
	8	L8W	0	3	3			
	9	P9L	0	5	5			
N-terminal domain	13	S13I	30	421	451			
	18	L18F	0	2	2			
	19	T19I	0	3	3			
	21	R21T	0	2	2			
	22	T22I	0	9	9			
	26	P26S	9	50	59			
	26	P26L	2	0	2			
	27	A27V	0	1	1			
	49	H49Y	0	6	6			
	67	A67V	2	20	22			
	69	H69Q	1	8	9			
	70	V70F	3	0	3			
	76	T76I	0	2	2			
	78	R78K	0	1	1			
	95	T95I	7	4	11			
	111	D111N	0	2	2			
	138	D138Y	1	2	3			
	141	L141F	0	1	1			
	142	G142-	0	1	1			
	143	V143F	0	1	1			
	143	V143-	0	1	1			
	152	W152C	30	396	426			
	153	M153I	1	4	5			
	155	S155N	0	6	6			
	173	Q173K	0	3	3			
	178	D178G	4	3	7			
	185	N185T	0	1	1			
	215	D215Y	7	1	8			
	222	A222V	0	1	1			
	235	I235T	0	1	1			
	242	L242P	0	1	1			
	243	A243-	0	1	1			
	251	P251S	1	1	2			
	252	G252V	0	3	3			
	254	S254F	1	0	1			
	257	G257S	0	3	3			
	258	W258L	0	19	19			
	261	G261C	0	1	1			
	261	G261V	0	1	1			
	Receptor binding domain	357	R357K	0	9	9		
367		V367I	7	3	10			
372		A372V	0	1	1			
449		Y449-	0	1	1			
452		L452R	30	421	451			
475		A475S	0	1	1			
477		S477I	0	1	1			
478		T478K	0	31	31			
484		E484K	0	1	1			
485		G485V	0	1	1			
494		S494P	0	6	6			
520		A520S	1	1	2			
Spike domain		565	F565L	0	1	1		
	574	D574Y	3	0	3			
	583	E583D	5	11	16			
	613	Q613H	7	5	12			
	614	D614G	191	795	986			
	622	V622I	0	1	1			
	640	S640F	0	2	2			
	657	N657D	1	0	1			
	Adjacent to S1/S2 junction	675	Q675H	2	6	8		
		677	Q677H	11	11	22		
		677	Q677P	0	11	11		
		679	N679K	0	1	1		
		681	P681H	0	34	34		
		681	P681R	1	0	1		
		684	A684V	0	1	1		
	688	A688V	5	8	13			
		701	A701S	0	1	1		
		716	T716I	0	1	1		
		719	T719I	0	3	3		
		732	T732A	7	62	69		
		769	G769V	3	5	8		
		771	A771D	0	2	2		
		779	Q779H	0	2	2		
		780	E780Q	0	1	1		
		792	P792L	0	1	1		
		812	P812S	0	1	1		
		813	S813G	0	2	2		
		Fusion peptide	818	I818V	0	1	1	
			845	A845S	0	4	4	
846	A846V		0	1	1			
Heptad repeat 1	879	A879S	2	0	2			
	936	D936H	3	1	4			
	936	D936Y	1	0	1			
	938	L938F	2	27	29			
	954	Q954K	0	2	2			
	1026	A1026S	1	0	1			
	1063	L1063F	0	1	1			
	1072	E1072K	0	10	10			
	1078	A1078S	0	3	3			
	1078	A1078V	0	1	1			
	1103	F1103L	1	0	1			
	1121	F1121V	1	0	1			
	1163	D1163V	0	2	2			
Heptad repeat 2	1168	D1168G	1	0	1			
	1171	G1171S	0	2	2			
	1176	V1176F	0	1	1			
	1191	K1191N	9	4	13			
	1195	E1195Q	0	1	1			
	1201	Q1201K	0	2	2			
Trans-membrane domain	1228	V1228L	0	1	1			
	1235	C1235F	5	5	10			
	1244	L1244F	1	0	1			
	1252	S1252P	1	0	1			
	1252	S1252F	0	1	1			

Table 5.S4 Individual characteristics of all persons tested (both positive and negative, including index case) living in one of the 319 households meeting inclusion criteria for household secondary attack rate analyses, stratified by strain classification of the household

	Non-West coast (N=571)	West coast			Total (N=1186)
		B.1.427 (N=325)	B.1.429 (N=261)	All West coast (N=615) ^a	
Sex					
Female	274 (48.0%)	143 (44.0%)	127 (48.7%)	287 (46.7%)	561 (47.3%)
Male	289 (50.6%)	177 (54.5%)	132 (50.6%)	321 (52.2%)	610 (51.4%)
Other	8 (1.4%)	5 (1.5%)	2 (0.8%)	7 (1.1%)	15 (1.3%)
Age Group					
Age ≤ 12	78 (13.7%)	31 (9.5%)	44 (16.9%)	79 (12.8%)	157 (13.2%)
Age 13-35	254 (44.5%)	129 (39.7%)	112 (42.9%)	257 (41.8%)	511 (43.1%)
Age 36-64	216 (37.8%)	142 (43.7%)	96 (36.8%)	246 (40.0%)	462 (39.0%)
Age ≥ 65	23 (4.0%)	23 (7.1%)	9 (3.4%)	33 (5.4%)	56 (4.7%)
Race/Ethnicity					
Hispanic/Latinx	511 (89.5%)	269 (82.8%)	228 (87.4%)	520 (84.6%)	1031 (86.9%)
Asian	17 (3.0%)	13 (4.0%)	12 (4.6%)	28 (4.6%)	45 (3.8%)
White/Caucasian	16 (2.8%)	15 (4.6%)	3 (1.1%)	18 (2.9%)	34 (2.9%)
Black or African American	4 (0.7%)	7 (2.2%)	3 (1.1%)	12 (2.0%)	16 (1.3%)
Other	23 (4.0%)	21 (6.5%)	15 (5.7%)	37 (6.0%)	60 (5.1%)
Occupation^b					
Food & Beverage	69 (12.1%)	49 (15.1%)	26 (10.0%)	80 (13.0%)	149 (12.6%)
Tradesperson	35 (6.1%)	19 (5.8%)	5 (1.9%)	24 (3.9%)	59 (5.0%)
Day laborer	19 (3.3%)	9 (2.8%)	9 (3.4%)	18 (2.9%)	37 (3.1%)
Healthcare	13 (2.3%)	9 (2.8%)	2 (0.8%)	11 (1.8%)	24 (2.0%)
Student	110 (19.3%)	42 (12.9%)	64 (24.5%)	110 (17.9%)	220 (18.5%)
Other	133 (23.3%)	79 (24.3%)	66 (25.3%)	149 (24.2%)	282 (23.8%)
No employment	86 (15.1%)	65 (20.0%)	44 (16.9%)	116 (18.9%)	202 (17.0%)
Day of Test Symptoms					
Asymptomatic	320 (56.0%)	191 (58.8%)	137 (52.5%)	349 (56.7%)	669 (56.4%)
Symptomatic	251 (44.0%)	134 (41.2%)	124 (47.5%)	266 (43.3%)	517 (43.6%)

^a29 individuals with S gene only sequence available.

^bOccupation information missing for 213 persons.

Table 5.S5 Adjusted attack rates from sensitivity analysis using Targeted Maximum Likelihood and Super Learning

	Adjusted	
	aRR (95% CI)	p-value
Class		
Non-West Coast	-	-
West Coast	1.17 (0.92-1.48)	0.190
Lineage		
B.1.427	1.13 (0.86-1.48)	0.387
B.1.429	1.30 (0.98-1.73)	0.071

5.10 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 BinaxNOW™ 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 cloudbased 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 MERSCoV. *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.

30. Crawford ED, Acosta I, Ahyong V, et al. Rapid deployment of SARS-CoV-2 testing: The CLIAHUB. *PLOS Pathog* 2020; 16:e1008966.
31. Quick J, Grubaugh ND, Pullan ST, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* 2017; 12:1261–1276.
32. R&d DP. COVID-19 ARTIC v3 Illumina library construction and sequencing protocol. 2020; Available at: <https://www.protocols.io/view/covid-19-artic-v3-illumina-libraryconstruction-an-bibt kann>. Accessed 23 February 2021.
33. Gohl DM, Garbe J, Grady P, et al. A rapid, cost-effective tailed amplicon method for sequencing SARS-CoV-2. *BMC Genomics* 2020; 21:863.
34. artic-network/artic-ncov2019. Available at: <https://github.com/artic-network/articncov2019>. Accessed 25 February 2021.
35. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinforma Oxf Engl* 2018; 34:3094–3100.
36. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinforma Oxf Engl* 2009; 25:2078–2079.
37. Grubaugh ND, Gangavarapu K, Quick J, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol* 2019; 20:8
38. chanzuckerberg/idseq-workflows. Chan Zuckerberg Initiative, 2021. Available at: <https://github.com/chanzuckerberg/idseq-workflows>. Accessed 25 February 2021.
39. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res* 2016; 44:D67-72.
40. 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; 5:1403–1407.

41. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013; 30:772–780.
42. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015; 32:268–274.
43. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 2018; 4:vey016
44. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinforma Oxf Engl* 1998; 14:817–818
45. Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol Biol Evol* 2013; 30:713–724.
46. Fraser C, Donnelly CA, Cauchemez S, et al. Pandemic potential of a strain of influenza A (H1N1): early findings. *Science* 2009; 324:1557–1561.
47. Ferretti L, Wymant C, Kendall M, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* 2020; 368.

6 SARS-COV-2 VARIANT EXPOSURES ELICIT ANTIBODY RESPONSES WITH DIFFERENTIAL CROSS-NEUTRALIZATION OF ESTABLISHED AND EMERGING STRAINS INCLUDING DELTA AND OMICRON

6.1 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.

6.2 INTRODUCTION

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 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.

6.3 METHODS

6.3.1 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

Table 6.S1.

6.3.2 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, Δ 242-244, 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, Δ 157-158, L452R, T478K, P681R, D614G, D950N), and B.1.1.529/omicron (32 spike mutations). All pseudovirus spike mutations are listed in **Table**

6.S2. 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).

6.3.2 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 x 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% CO₂ 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 (**Table 6.S3**) [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).

6.3.2 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 (NT50 and NT90) 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 NT50 and NT90 value to D614G pseudovirus NT50 and NT90 values in the same serum group. To compare neutralization titer across a panel of different pseudoviruses and serum groups, the log₂ fold-change compared to D614G pseudovirus was reported.

6.4 RESULTS

We compared NT50 and NT90 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 6.1**). Fold-changes in both NT50 and NT90 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 NT90, 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 6.2** and **Table 6.S3**). 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 NT90, 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.

6.5 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 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.

6.6 ACKNOWLEDGEMENTS

We thank Dr Chuka Didigu, Dorothy Park CRNA, Salu Ribeiro, and Bay Area Phlebotomy and Laboratory services for performing blood draws of study subjects; Dr Andreas Puschnik for providing the engineered cell line used in this study; Susana Elledge and Dr James Wells for providing reagents and advice on antibody assays; and Drs Peter Kim, Don Ganem, Sandy Schmidt, and Cori Bargmann for technical assistance and discussion.

This work was supported by the University of California San Francisco COVID fund (to J. D., M. L., J. L., and S. S.); the National Institutes of Health (grant numbers UM1AI069496 to D. H. and F31AI150007 to S. S.); the Chan Zuckerberg Biohub (to J. D. and D. H.); and the Chan Zuckerberg Initiative (to J. D. and D. H.).

J. D. is a member of the scientific advisory board of The Public Health Company, Inc and reports options granted for this service; and is scientific advisor for Allen & Co. All other authors

report no potential conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

6.7 FIGURES

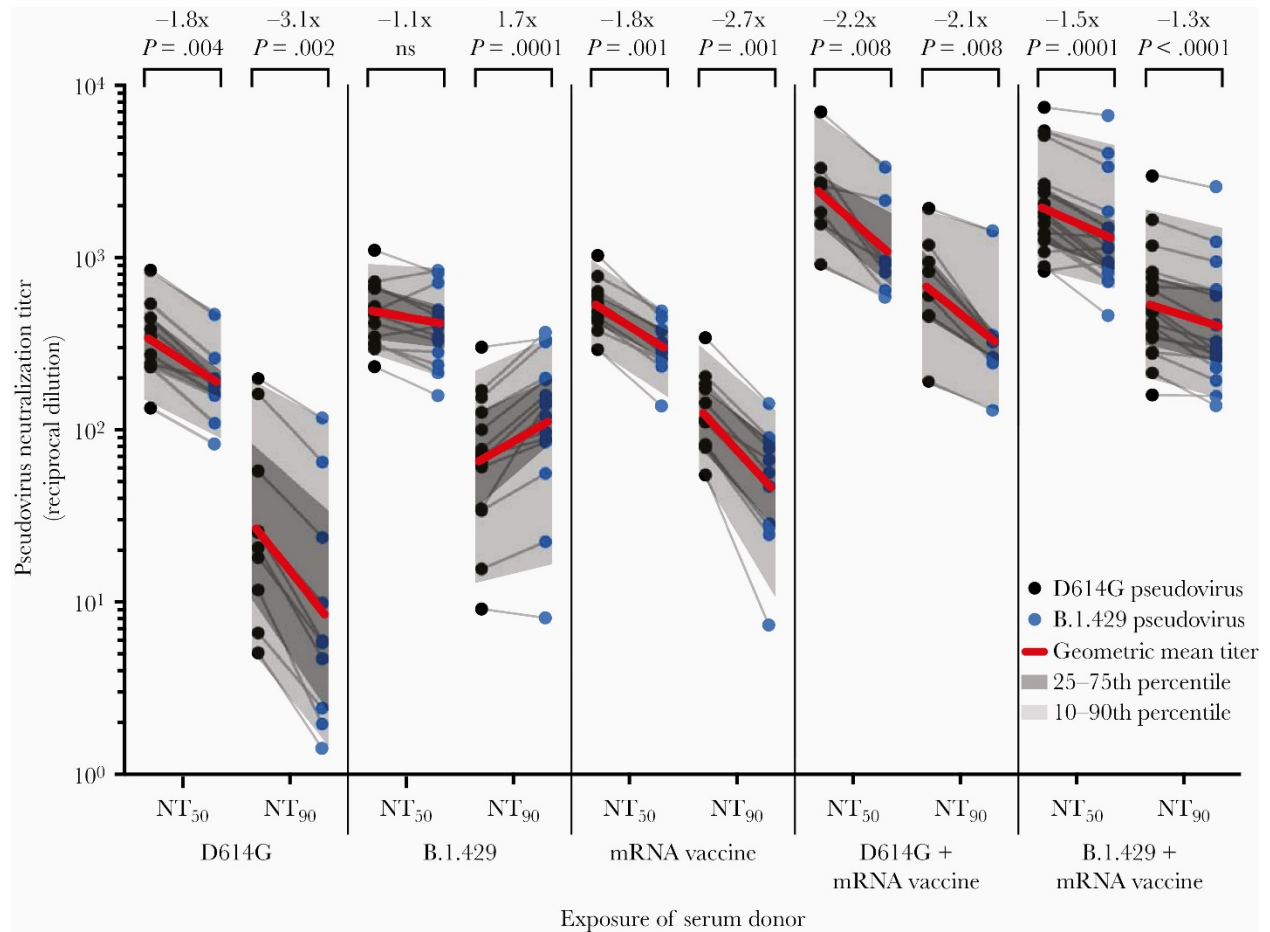


Figure 6.1 Neutralization of D614G and B.1.429 pseudoviruses by serum from individuals with different exposures

Plot of 50% and 90% pseudovirus neutralization titers (NT₅₀ and NT₉₀) 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 NT₅₀ and NT₉₀ 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.

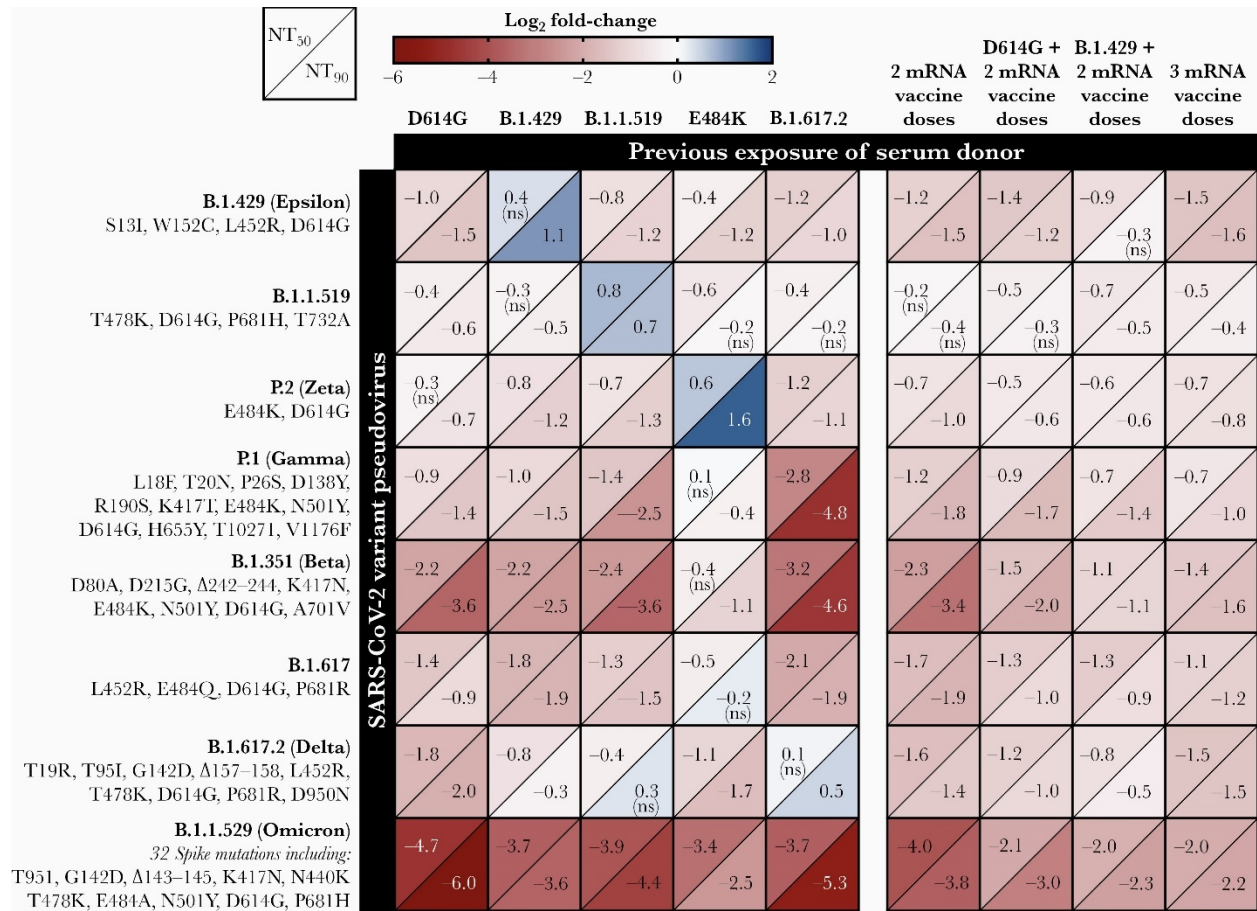


Figure 6.2 Change in variant pseudovirus neutralization titer relative to D614G

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 \log_2 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 \log_2 fold-change (blue) indicates an increase in neutralization titer for that variant relative to D614G pseudovirus, while negative \log_2 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.

6.8 TABLES

Table 6.S1 Serum sample metadata and exposure information

Serum group	Sample ID	Age	Sex	Symptoms (Yes/No/Unknown)	Days post positive COVID test	Days post most recent vaccine dose	Initial vaccine	Booster vaccine	Pango Lineage	GISAID Virus ID
B.1.429 (n=15)	Conv_1	54	Female	No	23				B.1.427	hCoV-19/USA/CA-UCSF-JD55/2021
	Conv_2	49	Female	No	35				B.1.427	hCoV-19/USA/CA-UCSF-JD423/2021
	Conv_3	42	Female	No	31				B.1.427	hCoV-19/USA/CA-UCSF-JD153/2021
	Conv_4	28	Female	No	42				B.1.427	hCoV-19/USA/CA-UCSF-JD612/2021
	Conv_5	43	Female	Yes	44				B.1.429	hCoV-19/USA/CA-UCSF-JD46/2021
	Conv_6	48	Male	No	46				B.1.429	hCoV-19/USA/CA-UCSF-JD818/2021
	Conv_7	41	Female	Yes	42				B.1.427	hCoV-19/USA/CA-UCSF-JD742/2021
	Conv_8	29	Male	Yes	41				B.1.429	hCoV-19/USA/CA-UCSF-JD883/2021
	Conv_9	58	Male	Yes	36				B.1.429	hCoV-19/USA/CA-UCSF-JD786/2021
	Conv_10	37	Female	Yes	43				B.1.429	hCoV-19/USA/CA-UCSF-JD866/2021
	Conv_11	46	Female	Yes	42				B.1.429	hCoV-19/USA/CA-UCSF-JD758/2021
	Conv_12	60	Female	Unknown	37				B.1.429	hCoV-19/USA/CA-UCSF-JD780/2021
	Conv_13	42	Male	Yes	45				B.1.427	hCoV-19/USA/CA-UCSF-JD796/2021
	Conv_14	21	Male	Yes	38				B.1.429	hCoV-19/USA/CA-UCSF-JD846/2021
	Conv_15	32	Female	Yes	36				B.1.429	hCoV-19/USA/CA-UCSF-JD775/2021
	Mean (SD)	42.0 (11.2)	33.3% Male 66.7% Female	80% Yes 20% No	38.7 (6.1)					
E484K (n=2)	Conv_16	31	Male	Yes	28				P.2	hCoV-19/USA/CA-UCSF-JD47/2021
	Conv_17	34	Male	Yes	44				B.1.526	hCoV-19/USA/CA-UCSF-JD776/2021
	Mean (SD)	32.5 (2.1)	100% Male 0% Female	100% Yes 0% No	36 (11.3)					
B.1.1.519 (n=6)	Conv_18	35	Female	No	28				B.1.1.519	hCoV-19/USA/CA-UCSF-JD73/2021
	Conv_19	40	Female	Yes	47				B.1.1.519	hCoV-19/USA/CA-UCSF-JD439/2021
	Conv_20	31	Male	Yes	37				B.1.1.519	hCoV-19/USA/CA-UCSF-JD172/2021
	Conv_21	51	Female	Yes	31				B.1.1.519	hCoV-19/USA/CA-UCSF-JD215/2021
	Conv_22	33	Female	No	51				B.1.1.519	hCoV-19/USA/CA-UCSF-JD355/2021
	Conv_23	35	Male	Yes	37				B.1.1.519	hCoV-19/USA/CA-UCSF-JD282/2021
	Mean (SD)	37.5 (7.3)	33.3% Male 66.7% Female	66.7% Yes 33.3% No	38.5 (8.9)					

Serum group	Sample ID	Age	Sex	Symptoms (Yes/No/Unknown)	Days post positive COVID test	Days post most recent vaccine dose	Initial vaccine	Booster vaccine	Pango Lineage	GISAID Virus ID
D614G (n=10)	Conv_24	57	Female	Yes	37			B.1.111	hCoV-19/USA/CA-UCSF-JD53/2021	
	Conv_25	62	Female	Yes	39			B.1.369	hCoV-19/USA/CA-UCSF-JD246/2021	
	Conv_26	30	Male	Yes	29			B.1.234	hCoV-19/USA/CA-UCSF-JD550/2021	
	Conv_27	55	Female	Yes	33			B.1.1.130	hCoV-19/USA/CA-UCSF-JD538/2021	
	Conv_28	38	Male	No	65			B.1.2	hCoV-19/USA/CA-UCSF-JD50/2021	
	Conv_29	42	Male	Yes	64			B.1.2	hCoV-19/USA/CA-UCSF-JD48/2021	
	Conv_30	43	Male	Yes	77			B.1.311	hCoV-19/USA/CA-UCSF-JD151/2021	
	Conv_31	64	Male	No	69			B.1.311	hCoV-19/USA/CA-UCSF-JD143/2021	
	Conv_32	62	Female	Yes	70			B.1.2	hCoV-19/USA/CA-UCSF-JD49/2021	
	Conv_33	35	Male	Yes	77			B.1.311	hCoV-19/USA/CA-UCSF-JD175/2021	
	Mean (SD)	48.8 (12.6)	50% Male 50% Female	80% Yes 20% No	56.0 (19.1)					
B.1.617.2 (n=3)	Conv_34	36	Female	Yes	47			B.1.617.2	hCoV-19/USA/CA-UCSF-JD1330/2021	
	Conv_35	42	Female	No	65			B.1.617.2	hCoV-19/USA/CA-UCSF-JD1373/2021	
	Conv_36	36	Female	Yes	49			B.1.617.2	hCoV-19/USA/CA-UCSF-JD1391/2021	
	Mean (SD)	38.0 (3.5)	0% Male 100% Female	66.7% Yes 33.3% No	53.7 (9.9)					
B.1.429 + mRNA vaccine (n=17)	Conv+vacc_1	47	Male	No	106	14	BNT162b2		hCoV-19/USA/CA-UCSF-JD434/2021	
	Conv+vacc_2	37	Male	No	104	20	BNT162b2		hCoV-19/USA/CA-UCSF-JD473/2021	
	Conv+vacc_3	42	Male	No	100	21	BNT162b2		hCoV-19/USA/CA-UCSF-JD363/2021	
	Conv+vacc_4	54	Female	Yes	110	32	BNT162b2		hCoV-19/USA/CA-UCSF-JD128/2021	
	Conv+vacc_5	35	Male	Unknown	94	33	BNT162b2		hCoV-19/USA/CA-UCSF-JD443/2021	
	Conv+vacc_6	39	Male	No	84	42	BNT162b2		hCoV-19/USA/CA-UCSF-JD694/2021	
	Conv+vacc_7	42	Female	No	105	46	BNT162b2		hCoV-19/USA/CA-UCSF-JD153/2021	
	Conv+vacc_8	33	Female	No	111	48	BNT162b2		hCoV-19/USA/CA-UCSF-JD516/2021	
	Conv+vacc_9	38	Female	No	96	50	BNT162b2		hCoV-19/USA/CA-UCSF-JD287/2021	
	Conv+vacc_10	55	Male	Yes	105	31	BNT162b2		hCoV-19/USA/CA-UCSF-JD138/2021	
	Conv+vacc_11	33	Female	Yes	121	16	BNT162b2		hCoV-19/USA/CA-UCSF-JD739/2021	
	Conv+vacc_12	40	Female	Yes	125	35	BNT162b2		hCoV-19/USA/CA-UCSF-JD757/2021	
	Conv+vacc_13	59	Female	Yes	92	14	BNT162b2		hCoV-19/USA/CA-UCSF-JD820/2021	
	Conv+vacc_14	37	Female	Yes	109	36	BNT162b2		hCoV-19/USA/CA-UCSF-JD315/2021	
	Conv+vacc_15	41	Male	Yes	128	14	BNT162b2		hCoV-19/USA/CA-UCSF-JD794/2021	
	Conv+vacc_16	36	Female	Yes	132	16	BNT162b2		hCoV-19/USA/CA-UCSF-JD788/2021	
	Conv+vacc_17	21	Male	Yes	98	14	BNT162b2		hCoV-19/USA/CA-UCSF-JD846/2021	
	Mean (SD)	40.5 (9.2)	47.1% Male 52.9% Female	52.9% Yes 41.2% No 5.9% Unknown	28.4 (13.1)	107.1 (13.2)				

Serum group	Sample ID	Age	Sex	Symptoms (Yes/No/Unknown)	Days post positive COVID test	Days post most recent vaccine dose	Initial vaccine	Booster vaccine	Pango Lineage	GISAID Virus ID	
D614G + mRNA vaccine (n=8)	Conv+vacc_18	42	Male	Yes	99	14	BNT162b2		B.1.2	hCoV-19/USA/CA/UCSF-JD48/2021	
	Conv+vacc_19	41	Male	No	106	22	BNT162b2		B.1.234	hCoV-19/USA/CA/UCSF-JD398/2021	
	Conv+vacc_20	32	Male	Yes	110	35	BNT162b2		B.1.2	hCoV-19/USA/CA/UCSF-JD380/2021	
	Conv+vacc_21	62	Female	Yes	114	21	BNT162b2		B.1.2	hCoV-19/USA/CA/UCSF-JD49/2021	
	Conv+vacc_22	64	Male	No	118	18	BNT162b2		B.1.311	hCoV-19/USA/CA/UCSF-JD143/2021	
	Conv+vacc_23	33	Male	Yes	134	14	BNT162b2		B.1.348	hCoV-19/USA/CA/UCSF-JD643/2021	
	Conv+vacc_24	40	Male	Yes	126	17	BNT162b2		B.1.2	hCoV-19/USA/CA/UCSF-JD303/2021	
	Conv+vacc_25	37	Male	Unknown	121	15	BNT162b2		B.1.311	hCoV-19/USA/CA/UCSF-JD358/2021	
	Mean (SD)	43.9 (12.3)	87.5% Male 12.5% Female	62.5% Yes 25.0% No 12.5% Unknown	19.5 (6.9)	116.0 (11.2)					
	2 dose mRNA vaccine (n=11)	Vacc_1	39	Male			35	mRNA-1273			
		Vacc_2	31	Male			36	mRNA-1273			
Vacc_3		22	Female			22	mRNA-1273				
Vacc_4		22	Female			27	mRNA-1273				
Vacc_5		30	Female			30	mRNA-1273				
Vacc_6		35	Female			60	mRNA-1273				
Vacc_7		26	Female			17	BNT162b2				
Vacc_8		24	Male			17	BNT162b2				
Vacc_9		30	Female			17	BNT162b2				
Vacc_10		30	Male			18	BNT162b2				
Vacc_11		52	Male			11	BNT162b2				
Mean (SD)	31.0 (8.7)	45.5% Male 54.5% Female			26.4 (13.8)						
3 dose mRNA vaccine (n=7)	Boost_1	37	Female			18	mRNA-1273	mRNA-1273			
	Boost_2	40	Male			14	mRNA-1273	mRNA-1273			
	Boost_3	47	Male			12	BNT162b2	BNT162b2			
	Boost_4	26	Female			43	BNT162b2	BNT162b2			
	Boost_5	59	Female			14	BNT162b2	mRNA-1273			
	Boost_6	41	Female			14	BNT162b2	mRNA-1273			
	Boost_7	31	Male			60	BNT162b2	BNT162b2			
Mean (SD)	40.1 (10.8)	42.9% Male 57.1% Female			25.0 (18.8)						

Table 6.S2 Spike gene mutations for each SARS-CoV-2 pseudovirus

Variant lineage	B.1	B.1.429	B.1.1.519	P.2	P.1	B.1.351	B.1.617	B.1.617.2	B.1.1.529
WHO variant name		Epsilon		Zeta	Gamma	Beta		Delta	Omicron
Spike mutations from Wuhan strain	D614G	S13I W152C L452R D614G	T478K D614G P681H T732A	E484K D614G	L18F T20N P26S D138Y R190S K417T E484K N501Y D614G H655Y T1027I V1176F	D80A D215G Δ 242-244 K417N E484K N501Y D614G A701V	L452R E484Q D614G P681R	T19R T95I G142D Δ 157-159 L452R T478K D614G P681R D950N	A67V Δ 69-70 T95I G142D Δ 143-145 Δ 211 L212I +214EPE G339D S371L S373P S375F K417N N440K G446S S477N T478K E484A Q493R N501Y Y505H T547K D614G H655Y N679K P681H N764K D796Y N856K Q954H N969K L981F

Table 6.S3 Table of measured pseudovirus neutralization titers in each serum group (mean of three replicates)

		Previous exposure of serum donor										WHO International Standard 20/150
		D614G	B.1.429	B.1.1.519	E484K	B.1.617.2	Vaccine	D614G + 2 vaccine doses	B.1.419 + 2 vaccine doses	3 vaccine doses		
D614G	NT50	466.33	715.85	857.14	238.18	788.49	971.23	1920.38	2110.10	1996.53	1024.31	
	NT90	58.33	108.88	210.96	26.39	107.74	226.17	610.77	531.97	283.29	232.49	
B.1.429 (Epsilon)	NT50	232.88	911.27	488.41	180.90	328.87	436.89	724.71	1138.98	726.70	510.75	
	NT90	21.24	232.90	94.13	11.55	52.61	78.07	264.12	431.40	94.32	87.54	
B.1.1.519	NT50	346.65	583.22	1451.35	156.24	596.06	851.35	1383.93	1335.42	1383.73	877.19	
	NT90	39.62	78.80	342.20	23.17	92.81	176.47	486.41	373.63	209.83	219.37	
P.2 (Zeta)	NT50	378.40	394.19	533.43	367.29	331.04	610.17	1327.03	1441.01	1195.40	396.78	
	NT90	35.77	46.28	87.56	80.75	50.08	112.48	398.71	344.75	160.47	46.46	
P.1 (Gamma)	NT50	258.73	371.78	319.53	246.71	127.26	411.55	1000.80	1320.49	1272.25	563.33	
	NT90	21.83	39.31	36.43	20.19	3.18	64.78	189.30	208.70	137.62	23.48	
B.1.351 (Beta)	NT50	101.94	157.34	160.93	176.67	80.48	192.19	671.21	993.56	771.19	271.48	
	NT90	4.77	19.87	17.59	12.00	2.59	22.00	149.89	251.30	96.25	25.12	
B.1.617	NT50	176.86	203.77	361.32	171.68	183.97	305.84	784.41	853.58	905.95	294.66	
	NT90	30.65	30.09	75.43	30.46	29.09	62.48	311.55	294.17	119.94	30.74	
B.1.617.2 (Delta)	NT50	135.88	419.90	634.49	112.62	851.44	310.92	822.25	1211.61	724.91	682.16	
	NT90	14.91	86.18	259.46	8.42	148.45	87.86	305.87	386.89	102.22	89.97	
B.1.1.529 (Omicron)	NT50	17.57	53.47	55.83	22.38	59.48	61.20	445.04	545.99	497.75	77.26	
	NT90	0.90	8.75	9.90	4.79	2.68	16.93	79.08	107.89	62.90	15.82	

6.9 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 communitybased 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.

7 SARS-COV-2 TRANSMISSION DYNAMICS AND IMMUNE RESPONSES IN A HOUSEHOLD OF VACCINATED PERSONS

7.1 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.

7.2 INTRODUCTION

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.

7.3 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 tocilizumab 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 7.1A** and **Table 7.S1**.

7.4 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 5.S1, BioProject PRJNA790937). Despite incomplete recovery, the partial sequence from Individual 1 (17%) contained mutations consistent with the Gamma variant (**Table 7.S2**). 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 pseudovirus neutralization assay [7]. Sera from members of this household demonstrated a wide range of neutralization (**Figure 7.1B**). Individual 1 had a much lower neutralizing antibody titer compared with the fully vaccinated individuals (D614G 50% neutralization titer [NT50] = 4.4x lower, Gamma NT50 = 6.3x 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 NT50 = 4.5x higher, Gamma NT50 = 5.0x 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 NT50 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, 8].

Additionally, we tested for anti-interferon (IFN)- α 2 autoantibodies, a marker correlated with severe COVID-19 and poor patient outcomes [9, 10]. Using serum from patients with autoimmune polyglandular syndrome type 1 (APS1), an autoimmune syndrome where patients frequently develop an abundance of anti-IFN- α 2 antibodies, as a benchmark for verified IFN autoimmunity, we measured for anti-IFN- α 2 antibody presence using a radioligand binding assay (RLBA) [9]. Serum from Individual 1, who had the most severe response to infection, exhibited positive anti-IFN- α 2 antibody signal while the other family members had negative titers (Figure 5.1C).

7.5 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- α 2 autoantibodies revealed that serum from Individual 1 contained high levels of antibodies against IFN- α 2, a trait enriched among patients with life-threatening COVID-19 pneumonia [10]. 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 [10]. 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.

7.6 ACKNOWLEDGEMENTS

We thank Dr. Chuka Didigu; Dorothy Park, CRNA; Salu Ribeiro; and Bay Area Phlebotomy and Laboratory Services for performing blood draws of study subjects. We thank Dr. Andreas Puschnik for providing the engineered cell line used in this study. We thank Drs. Peter Kim, Don Ganem, Sandy Schmidt, and Cori Bargmann for technical assistance and discussion.

This work was supported by the University of California San Francisco COVID fund (to J. D., M. L., J. L., and S. S.), the National Institutes of Health (grant number UM1AI069496 to D. H. and grant number F31AI150007 to S. S.), the Division of Intramural Research of the National Institute of Allergy and Infectious Diseases (NIAID) (ZIA number AI001175 to M. S. L.), the Chan Zuckerberg Biohub (to J. D. and D. H.), and the Chan Zuckerberg Initiative (to J. D. and D. H.). Potential conflicts of interest. J. D. is a member of the scientific advisory board of The Public Health Company, Inc, and is a scientific advisor for Allen & Co. J. D. also reports stock options granted for service on the Scientific Advisory Board of The Public Health Company; reports payment or honoraria for various small invited academic lectures at university, approximately 10 of these over the past 36 months; and is a member of the board of the Chan Zuckerberg Biohub, a nonprofit 501c3 scientific research organization affiliated with University of California San Francisco (UCSF), Stanford, and University of California Berkeley. None of the other authors have any potential conflicts. C. M. reports grants or contracts from the National Institutes of Health (NIH), Stupski Foundation, and J. P. McGovern Foundation paid to the institution outside of the submitted work, and Chan Zuckerberg Biohub honoraria for panel discussion on vaccine hesitancy. D. H. reports grants or contracts from NIH outside of the submitted work. M. A. reports grants or contracts NIH/NIAID R37AI097457 (NIH grant to UCSF);

consulting fees to self from Abolieris, Inc, Sana, Inc, Rubius, Inc, and NGM Bio, Inc; is a member of the scientific advisory board for Imcyse, Inc; is President, Federation of Clinical Immunology Societies (FOCIS), which is a not-for-profit immunology society; and owns stock in Medtronic, Inc, and Merck, Inc. All other authors report no potential conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

7.7 SUPPLEMENTARY METHODS

7.7.1 Specimen collection and processing

Remnant clinical samples were collected from Zuckerberg San Francisco General Hospital (ZSFG) under a research study approved by the Committee on Human Subjects Research (CHR) at the University of California, San Francisco, UCSF. Nasal swab samples collected in viral transport media were first inactivated at 65 C for 30 minutes prior to 1:1 dilution in 2x DNA/RNA Shield. Nasal swab samples and sera from the other household members were obtained under informed consent through “Unidos en Salud,” a community-based SARS-CoV-2 testing, surveillance, and vaccination program, as previously described [13]. Nasal swabs were inactivated in DNA/RNA shield (Zymo). Serum samples were inactivated at 56C for 30 minutes prior to experimentation.

Additional sera from healthy recipients of BNT162b2/Pfizer and mRNA-1273/Moderna vaccines were collected as previously described [8]. Plasma samples from healthy pre-COVID individuals were obtained from the New York Blood Center, under informed consent research protocols. Autoimmune Polyglandular Syndrome Type 1 (APS1)-positive control samples were collected as previously described [11]. All APS1 patients were enrolled in research study

protocols approved by the NIAID, NIH Clinical Center, and NCI Institutional Review Board Committee. Participants provided written informed consent for study participation. All NIH patients gave consent for passive use of their medical record for research purposes.

7.7.2 Whole viral genome sequencing

RNA was extracted and processed for qRT-PCR and genome recovery as previously described [14]. Viral genomes were recovered using ARTIC Network V3 primers [15] and sequenced on a Nanopore MinION. Consensus genomes were generated by aligning raw fastq files to the Wuhan-Hu-1 reference sequence (GenBank: MN908947.3) using CZ-ID (former known as IDseq) [16]. The CZ-ID pipeline uses the ARTIC Network's nCoV-2019 novel coronavirus Nanopore bioinformatics protocol [17] with the following parameter changes: normalise set to 1000; medaka model set to r941_min_high_g360; and min-length set to 350. The GitHub repos used were ARTIC Nanopore protocol (<https://github.com/artic-network/artic-ncov2019.git>) and the ARTIC field bioinformatics pipeline (<https://github.com/artic-network/fieldbioinformatics>). The exact command lines used were:

```
artic guppyplex --skip-quality-check --min-length 350 --max-length 1500 --directory [directory] --prefix run_filtered
```

```
artic minion --medaka --medaka-model r941_min_high_g360 --normalise 1000 --threads 8 --strict --scheme-directory [directory]/artic-ncov2019/primer_schemes --read-file run_filtered_barcode02.fastq nCoV-2019/V1200 [name]
```

Recovered genomes with >90% coverage were deposited in GISAID [18]. Raw sequencing data can be accessed with BioProject ID: PRJNA790937.

7.7.3 Pseudovirus neutralization assay

Neutralization assays with pseudotyped recombinant vesicular stomatitis virus expressing GFP in place of the glycoprotein (rVSV Δ G-GFP) were performed as previously described [7]. Briefly, a CMV-driven expression vector was used to produce SARS-CoV-2 pseudovirus bearing spike protein of either the D614 wild-type strain or the Gamma variant. Pseudovirus titering and neutralization assays were performed on Huh7.5.1 cells overexpressing ACE2 and TMPRSS2 by flow cytometry to measure percent GFP positivity [19]. Mean serum neutralizing antibody titers (NT50) were calculated as an average of three independent experiments each with technical duplicates. Cells were verified to be free of mycoplasma contamination with the MycoAlert Mycoplasma detection kit (Lonza).

7.7.4 Anti-Type I Interferon Alpha 2 (anti-IFN- α 2) Radioligand Binding Assay (RLBA)

Radioactive counts per minute (cpm) were measured by immunoprecipitation of full-length Type I Interferon Alpha 2 (IFN- α 2) protein with a FLAG-Myc tag with either 2.5 μ L of plasma or 1 μ L anti-myc positive control antibody as previously described [9,20]. Antibody index is defined as the (sample cpm value–mean blank cpm value)/(positive control antibody cpm value–mean blank cpm value).

7.8 FIGURES

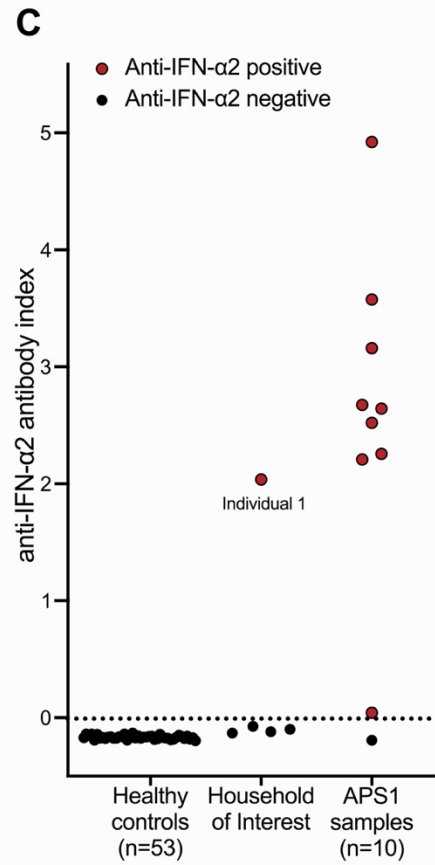
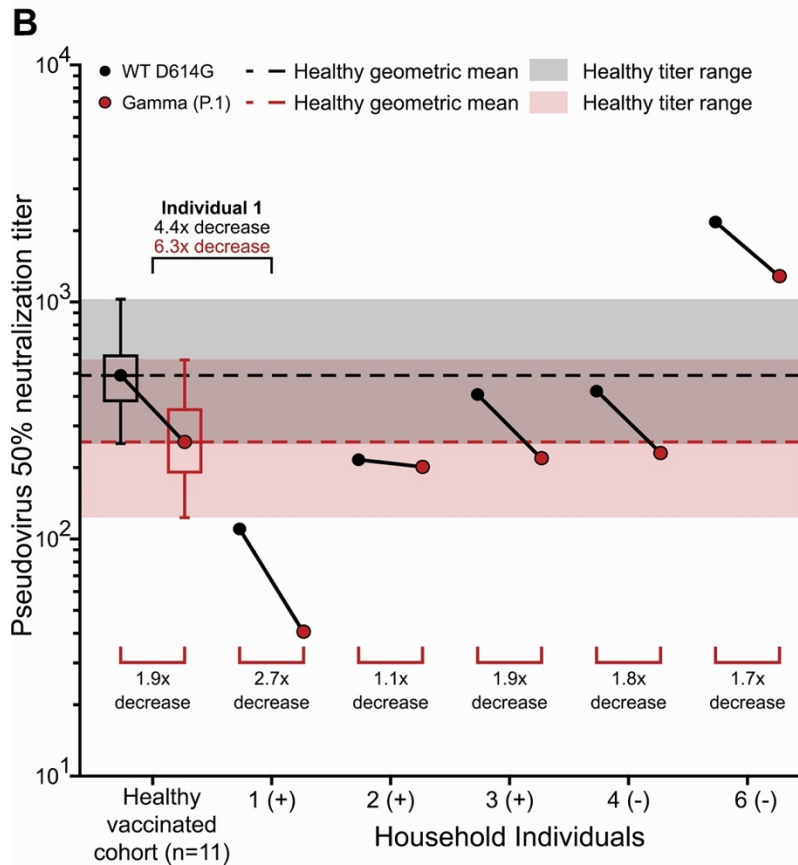
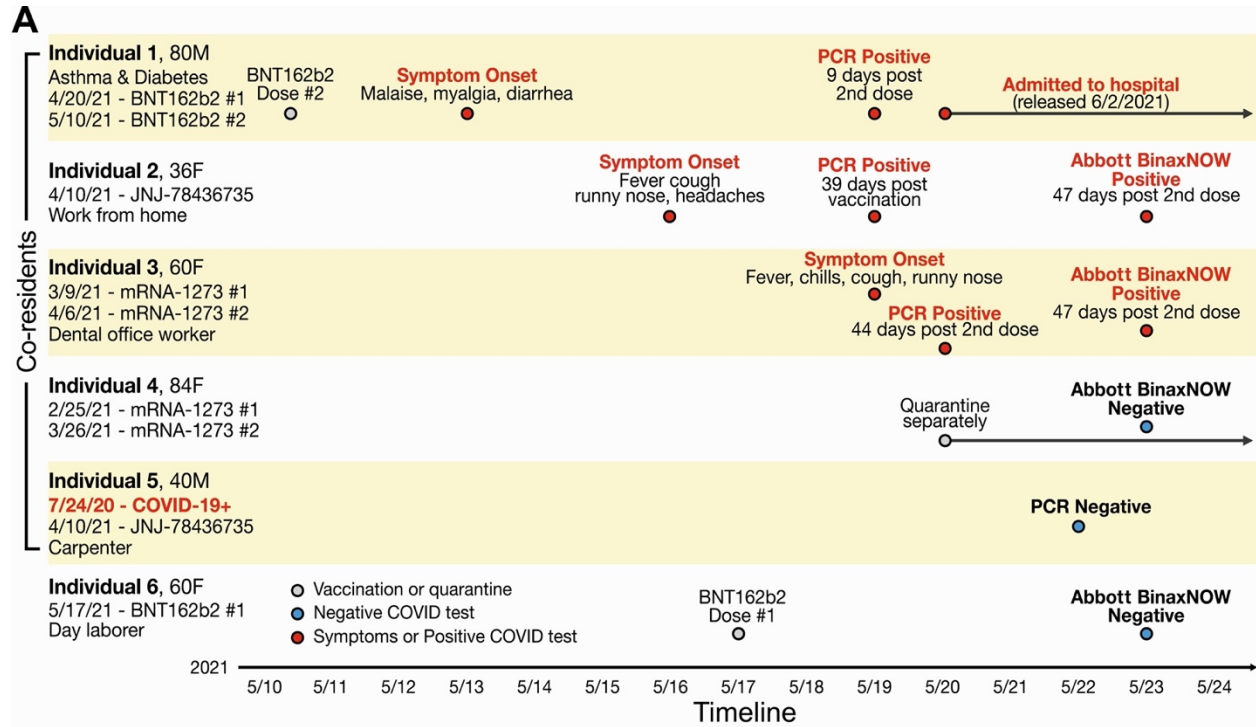


Figure 7.1 Serum samples from household individuals reveal diverse neutralization capabilities as well as presence of anti-IFN- α 2 auto-antibodies in Individual 1

(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 S1**. (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 S3**). For the healthy vaccinated donor cohort, geometric mean titer (dashed lines), interquartile range (boxes), and full range (shaded region) are shown for D614G (black) 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- α 2 autoantibodies are absent from all assayed pre-pandemic healthy controls ($n = 42$) and vaccinated healthy controls ($n = 11$) [8]. In this household, only Individual 1 demonstrated the presence of anti-IFN- α 2 auto-antibodies. Autoimmune polyglandular syndrome type 1 (APS1) patient sera are used as positive controls [11]; 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.

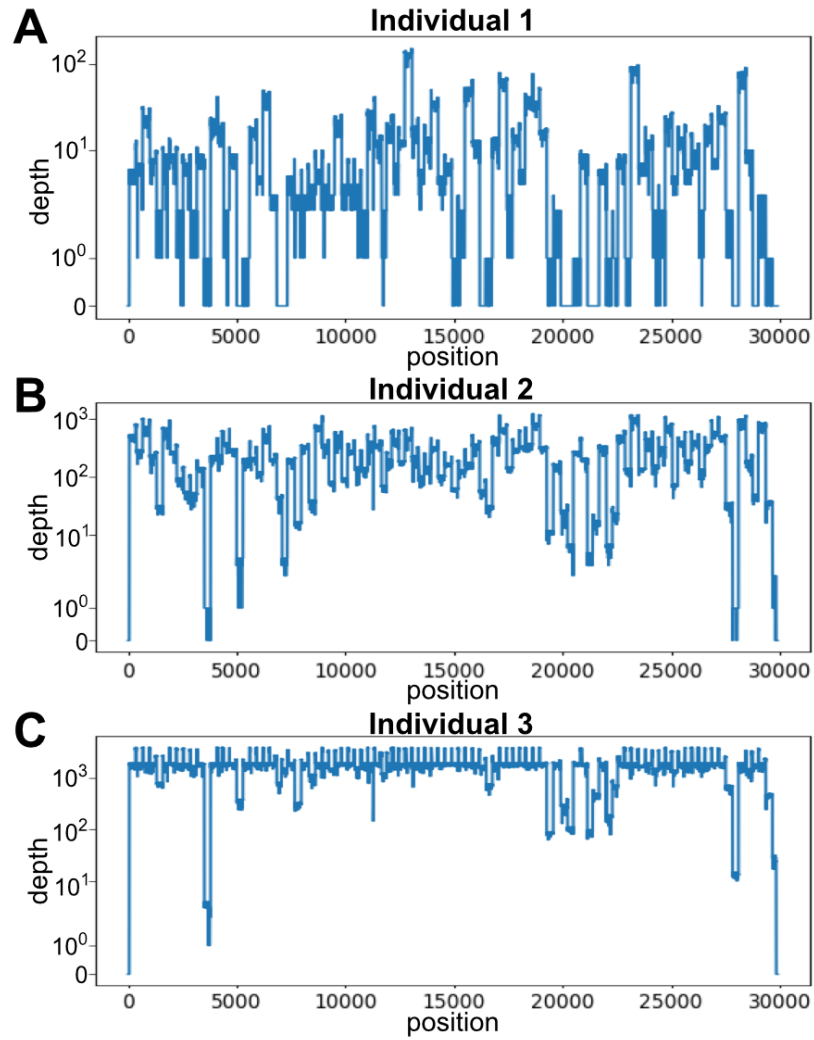


Figure 7.S1 Sequencing depth and coverage of recovered SARS-CoV-2 genomes.
Genome depth and coverage of SARS-CoV-2 Gamma variant recovered from (A) Individual 1, (B) Individual 2, and (C) Individual 3.

7.9 TABLES

Table 7.S1 Study household description and metadata

	Individual 1	Individual 2	Individual 3	Individual 4	Individual 5	Individual 6
Age	80	36	60	84	40	60
Sex	Male	Female	Female	Female	Male	Female
Co-morbidities	Diabetes, Asthma	N/A	N/A	N/A	N/A	N/A
Vaccine	BNT162b2	JNJ-78436735	mRNA-1273	mRNA-1273	JNJ-78436735	BNT162b2
Dose 1 Date	4/20/21	4/10/21	3/9/21	2/25/21	4/10/21	5/17/21
Dose 2 Date	5/10/21	N/A	4/6/21	3/26/21	N/A	TBD
Symptom onset	Yes	Yes	Yes	No	No	No
Date	5/13/2021 3 days post-second dose	5/16/2021 36 days post-vaccination	5/19/2021 43 days post-second dose			
Symptoms	malaise, myalgia, and diarrhea	fever, cough, rhinorrhea, headaches	fever, chills, cough, and rhinorrhea.			
COVID Test Results						
	5/19/2021 9 days post-second dose PCR Positive	5/19/2021 39 days post-vaccination PCR Positive	5/20/2021 44 days post-second dose PCR Positive	5/23/2021 58 days post-second dose BinaxNOW Negative	7/20/2020 Before vaccination PCR Positive	7/20/2020 Before vaccination PCR Negative
		5/23/2021 43 days post-vaccination BinaxNOW Positive	5/23/2021 47 days post-second dose BinaxNOW Positive		5/22/2021 42 days post-vaccination PCR Negative	5/23/2021 6 days post-second dose BinaxNOW Negative
Hospitalization	5/20/2021 - 6/2/2021	N/A	N/A	N/A	N/A	N/A
Sample Collection						
Nasal Swab	5/21/21	5/23/21	5/23/21	5/23/21	N/A	5/23/21
Serum	5/27/21	5/23/21	5/23/21	5/23/21	N/A	5/23/21

Table 7.S2 Characteristic mutations of the Gamma variant in SARS-CoV-2 positive individuals.

	Individual 1	Individual 2	Individual 3	P.1 Characteristic Mutation
BioProject PRJNA790937	SRR17281221	SRR17281220	SRR17281219	
Orf1a				
ORF1a:S1188L	yes	yes	yes	yes
ORF1a:K1795Q	Undetermined	yes	yes	yes
ORF1a:S3675-	yes	yes	yes	yes
ORF1a:G3676-	yes	yes	yes	yes
ORF1a:F3677-	yes	yes	yes	yes
ORF1b				
ORF1b:P314L	Undetermined	yes	yes	yes
ORF1b:A1219S	no	No	yes	no
ORF1b:E1264D	yes	yes	yes	yes
ORF1b:V2073L	Undetermined	yes	yes	no
S				
S:L18F	Undetermined	Undetermined	yes	yes
S:T20N	Undetermined	yes	yes	yes
S:P26S	Undetermined	yes	yes	yes
S:D138Y	Undetermined	yes	yes	yes
S:R190S	Undetermined	yes	yes	yes
S:K417T*	Undetermined	yes	yes	yes
S:E484K*	Undetermined	yes	yes	yes
S:N501Y*	Undetermined	yes	yes	yes
S:D614G	yes	yes	yes	yes
S:H655Y	Undetermined	yes	yes	yes
S:T1027I	Undetermined	yes	yes	yes
S:V1176F	Undetermined	yes	yes	yes
S:R346K*	Undetermined	no	no	no
S:L452R*	Undetermined	no	no	no
S:T478K*	Undetermined	no	no	no
*Substitutions of Concern for SARS-CoV-2 Monoclonal Antibody Therapies				
ORF3a				
ORF3a:S253P	Undetermined	yes	yes	yes
ORF8				
ORF8:E92K	yes	yes	yes	yes
ORF9b				
ORF9b:Q77E	Undetermined	yes	yes	no
N				
N:P80R	Undetermined	yes	yes	yes
N:R203K	Undetermined	yes	yes	yes
N:G204R	Undetermined	yes	yes	yes

Table 7.S3 Healthy vaccinated cohort metadata

Healthy Vaccinated Cohort (n=11)	
Vaccine received	
BNT162b2/Pfizer	5 (45.5%)
mRNA-1273/Moderna	6 (54.5%)
Age	
	31 ± 8.7
Sex	
Female	6 (54.5%)
Male	5 (45.5%)
Days post-second dose	
	26.4 ± 13.8

7.10 REFERENCES

1. COVID-19 Dashboard. Available at: <https://coronavirus.jhu.edu/map.html>. Accessed 11 September 2021.
2. Ritchie H, Mathieu E, Rod s-Guirao L, et al. Our World in Data: Coronavirus Pandemic (COVID-19). Our World in Data 2020; Available at: <https://ourworldindata.org/covidvaccinations>. 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; 70:1088–1093.
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; 592:616–622.
6. Vignier N, Brot 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; 27:2673–2676.
7. Hoffmann M, Kleine-Weber H, P hlmann 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.
8. 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.

9. 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.
10. Bastard P, Rosen LB, Zhang Q, et al. Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science* 2020; 370:eabd4585.
11. 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.
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.
13. 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. *Clin Infect Dis* 2021; :ciab283.
14. Crawford ED, Acosta I, Ahyong V, et al. Rapid deployment of SARS-CoV-2 testing: The CLIAHUB. *PLoS Pathog* 2020; 16:e1008966.
15. Quick J. nCoV-2019 sequencing protocol v2 (GunIt). 2020; Available at: <https://www.protocols.io/view/ncov-2019-sequencing-protocol-v2-bdp7i5rn>. Accessed 11 September 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:giaa111.
17. Loman N, Rowe W, Rambaut A. nCoV-2019 novel coronavirus bioinformatics protocol. 2020; Available at: <https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>. Accessed 29 December 2021.

18. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* 2017; 22:30494.
19. 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.
20. Vazquez SE, Bastard P, Kelly K, et al. Neutralizing Autoantibodies to Type I Interferons in COVID-19 Convalescent Donor Plasma. *J Clin Immunol* 2021; 41:1169–1171.

8 LOW-COST TOUCHSCREEN DRIVEN PROGRAMMABLE DUAL SYRINGE PUMP FOR LIFE SCIENCE APPLICATIONS

8.1 ABSTRACT

Syringe pumps are powerful tools able to automate routine laboratory practices that otherwise consume large amounts of manual labor time. Commercially available syringe pumps are expensive, difficult to customize, and often preset for a narrow range of operations. Here, we show how to build a programmable dual syringe pump (PDSP) that overcomes these limitations. The PDSP is driven by a Raspberry Pi paired with a stepper motor controller to allow maximal customization via Python scripting. The entire setup can be controlled by a touchscreen for use without a keyboard or mouse. Furthermore, the PDSP is structured around 3D printed parts, enabling users to change any component for their specific application. We demonstrate one application of the PDSP by using it to generate whole cell lysates using a cell homogenizer in an automated fashion

8.2 HARDWARE IN CONTEXT

Syringe pumps have a wide variety of uses across fields from engineering to biology. Their primary purpose is to continuously dispense precise volumes over a set amount of time. They save time by running unsupervised and provide more consistency than human hands. A dual syringe pump allows for two syringes to have coordinated actions, broadening the potential applications. In particular, dual syringe pumps have the power to automate a number of routine and repetitive protocols in the life sciences.

Our lab initially conceived of building a dual syringe pump that could be used to make *Plasmodium falciparum* whole cell extracts used for in vitro translation. Previously, we generated lysates by passing purified infected red blood cells through an Isobiotech cell homogenizer using a ball bearing with 4 μ m clearance. Frozen pellets of purified parasites were thawed, loaded into one 3 mL syringe, and passed through the homogenizer into a second 3 mL syringe. The lysate was then passed back into the first syringe. This back and forth cycle was repeated up to 20 times per parasite pellet [1]. This process takes between 20 and 30 min per parasite pellet and the resistance in the device makes it physically taxing on the wrists and thumbs. In the lab, lysate generation was often a rate-limiting step when making in vitro translation extracts due to the manual and tedious nature of the process. It also entailed unacceptable amounts of user-dependent variation between lysate preparations. In the worst cases, syringes would break or the plunger would deform as pressure was applied unevenly. A customizable PDSP addresses all of these issues not only for our uses, but also for other applications of the Isobiotech Cell homogenizer such as *C. elegans* lysis [2] and mammalian cell culture homogenization [3].

We reasoned that this process could be easily replaced with a programmable dual syringe pump. However, commercially available dual syringe pumps have a number of limitations, including cost and flexibility. As of this writing, commercially available dual syringe pumps cost upwards of \$1500 [4] and coordinated programmable motion tends to be limited.

Additionally, the physical configuration of commercial products makes alternative mounting options difficult. Previously described lab-built syringe pumps include single syringe pumps [5,6], or pumps that operate as part of a larger device, such as an auto-sampler [7], a 3D printer [8] or a pH-stat device [9]. To our knowledge, there are no current solutions that feature a customizable graphical user interface (GUI) touchscreen, which greatly simplify use, eliminating the need for keyboard and/or mouse. Thus, we designed, built, and tested a PDSP that can be easily used in the lab using a touch-screen but that can also be customized for any use. This PDSP drives two independent pumps operating under a single microcontroller which makes it not only amenable to synchronized pumping patterns, but also makes it extensible (up to 16 pumps) for multiplexed liquid handling operations.

8.3 HARDWARE DESCRIPTION

Our custom PDSP is constructed on an extruded aluminum frame that can be mounted horizontally or vertically. For our specific application, we chose a vertical mount to allow the cell homogenizer to be immersed in ice (**Figure 8.1A-B**). The PDSP utilizes two NEMA-14 stepper motors (StepperOnline) that create precise volume changes even at high torque. The motors are driven by a Pi-Plate MOTORplate controller (Pi-Plates, Inc.) mated to a Raspberry Pi (v3 Model B). Integrated limit switches provide for simple and accurate “homing” procedures. To make the PDSP easy to use for routine laboratory stand-alone use, the PDSP is operated via an attached touchscreen (Landzo), without a keyboard or mouse.

While this PDSP is specialized to lysate generation, alternative applications may have different requirements. We designed this device so that it could be adapted to many different environments, such as a biohazard hood, and for many different tasks, such as microfluidic experiments [5]. To allow for maximum customization, we made the PDSP modular with 3D printed parts that can be interchanged to accommodate different volume syringes. By simply

changing the dimensions of the printed parts the syringes can be set to any distance apart. The cell homogenizer holder can be interchanged with any other user-designed holder. Our custom Python/Tkinter interface can be easily configured to drive the stepper motors at different speeds allowing for a range of flow rates or even gradients of flow rates. The Pi-Plate MOTORplates can be stacked, allowing a single Pi and interface to simultaneously control up to 8 PDSPs (16 syringes) for high volume production environments.

8.4 DESIGN FILES

Table 8.1 contains links to the STL design files for all 3D printed parts described here:

- The Motor Base (**Figure 8.2A**) secures the two stepper motors and holds the Aluminum Heat Sink Plate against the T-profile rails.
- The Toe Hold (**Figure 8.2B**) positions the linear motion shafts against the T-profile rails.
- The Syringe Stabilizer (**Figure 8.2C**) holds the syringe body in place by securing both the barrel and the barrel flange parallel to the movement direction of the plunger.
- The left and right Plunger Movers (**Figure 8.2D**) hold the plunger flange of the syringes and move them along the linear motion shaft as the T8 threaded rods turn according to the stepper motors to change the volume.
- The Pi Base (**Figure 8.2E**) and Cover (**Figure 8.2F**) secure the Raspberry Pi along the T-profile rails and insulates it from short-circuiting against nearby conductive material. It also protects the Raspberry Pi from dust and accidental splashes from the ice bucket.
- The Screen Mount Top and Bottom (**Figure 8.2G**) hold the touch screen at eye-level when the PDSP is constructed vertically. Situated in front of the motors, it keeps the touchscreen a safe distance from the generated heat.

- The Cell Homogenizer Holder (**Figure 8.2H**) positions the cell homogenizer in alignment with luer lock syringes such that no additional tubing is necessary. It also allows the cell homogenizer to be in contact with ice at all times.
- The Right-Angle Brackets are used to position the syringe pump vertically on a breadboard. If desired, they can be 3D printed rather than purchased.
- The Aluminum Heat Sink Plate is placed between the stepper motors on the Motor Base and the T-profile rails to rapidly disperse heat generated from the stepper.

8.5 BILL OF MATERIALS

Table 8.2 contains all purchased items with component part numbers, required quantities, and cost.

8.6 BUILD INSTRUCTIONS

The essential PDSP components are 3D printed. The remaining hardware can be acquired online from common hardware suppliers, such as McMaster-Carr. All electronic parts can be acquired from common electronics suppliers, such as Adafruit or Microcenter (see Section 8.5).

The additional tools required for assembly include:

- M2.5 Tap
- M3 Tap
- Metric Hex Key Set
- Imperial Hex Key Set
- Philips Screwdriver
- Wire Stripper
- Soldering iron

- Hack-saw or chop-saw
- Drill Press

8.6.1 Preparation of electronic components

It is useful to prepare the electronic components prior to assembly. First, solder wires to the GND pin and the N pin on the leaf switches as outlined in **Figure 8.3**. Wrap the exposed solder and pin with heat shrink to protect the connection. To keep the electronic wiring neat, braid the four wires of each stepper motor and twist together the two wires of each leaf switches. Next, attach one heat sink onto the rear of each stepper motor by applying an appropriately sized piece of thermal tape to the rear end-cap and firmly pressing the heat sink to the tape. To prepare the motor power supply, cut and strip the wires of the 2.1 mm female/male barrel jack extension cable to separate the power and ground wires.

Finally, attach the Raspberry Pi onto the Pi base (**Figure 8.2E**) using four M2.5 x 5mm screws. Next, attach the PiPlates Motor Plate to the Raspberry Pi header pins by carefully applying even pressure on the plate while pushing the two components together to prevent bending any pins. Once attached, secure the PiPlates Motor Plate onto the Pi base using another four M2.5 x 5mm screws.

8.6.2 Hardware preparation

3D print each of the necessary parts from the provided design files, including four right-angle brackets if not using ThorLab's precision cut components. We recommend printing with ABS material with a low-density fill. Cut the two T8 lead screws to 25 cm each and the two linear motion shafts to 21.5 cm each. Saw the aluminum plate to the appropriate dimensions according to the Aluminum Heat Sink Plate STL file (**Supplementary File 1** available with the original publication). Using a drill press, drill two through holes as specified into the aluminum plate for later assembly onto the T-profile rails. Tap M2.5 and M3 holes into the 3D printed parts as indicated in their respective STL files.

8.6.3 Hardware assembly

The complete assembly of the PDSP as described below can be seen as a time-lapse in **Supplementary File 2**, which is available with the original publication.

1. Attach the compact end-feed fastener with M5 x 5mm screws to each through hole of the Motor Base (**Figure 8.2A**), Toe Hold (**Figure 8.2B**), Syringe Stabilizer (**Figure 8.2C**), Pi Base (**Figure 8.2E**), and Cell Homogenizer Holder (**Figure 8.2H**). Attach the compact end-fasteners of two M5 x 8mm through the previously drilled holes aluminum Heat Sink Plate (**Figure 8.4A**).
2. Attach each of the two pillow block bearings to the Extract Maker Holder (**Figure 8.4B**) using M3 x 20mm socket head screws.
3. Insert one copper nut and one linear ball bearing into each of the Plunger Movers (**Figure 8.2D**). Attach the copper nut using four M2.5 x 5mm socket head screws per nut and secure the linear ball bearing into its designated slot with silicone if necessary (**Figure 8.4C**).
4. Thread the one T8 lead screws into the copper nut and slide one linear motion shaft through the linear ball bearing on each of the two Plunger Movers (**Figure 8.4C**).
5. Making sure each part is oriented correctly, slide the four metal bars through the Syringe Stabilizer, and secure into the Cell Homogenizer Holder. The arms on each of the plunger mover should be on the outside of the linear structure and each pointed towards the Syringe Stabilizer (**Figure 8.4B** and **8.4D**).
6. On the other end, slide the four metal bars towards the Toe Hold making sure the linear motion shaft is secured by the Cell Homogenizer Holder on one end and by the Toe Hold on the other end (**Figure 8.4D**).
7. Slide the Toe Hold, the Syringe Holder, and the Cell Homogenizer Holder together onto the T-profile rails by correctly slotting the compact end-feeder into the rails. Do not tighten the M5 screws yet (**Figure 8.4D**).

8. Attach and secure the stepper motors onto the motor base using four M3 x 5mm per motor (**Figure 8.4E**).
9. Slide the motor base with attached motors onto the T-profile rails on the side closest to the Toe Hold. Do not yet tighten the M5 screws.
10. Attach the couplers between the T8 threaded rods and the stepper motors by carefully sliding all the moving parts along the T-profile rails towards each other. After making sure that the T8 threaded rods and the motors are completely aligned, secure the couplers by tightening the attached screws (**Figure 8.4F**).
11. Slide the Aluminum Heat Sink Plate onto the T-profile rails beneath the stepper motors on the Motor Base.
12. If desired, ensure there is sufficient space between the Cell Homogenizer Holder and the end of the T-profile rails to place an ice bucket. Finally, secure all the parts to the T-profile rails by tightening all ten of the M5 screws.
13. Attach the Screen Mount-Bottom (**Figure 8.2I**) onto the Motor Base using four M2.5 x 20mm screws. The correct orientation should allow the touchscreen to rest on the base without bending the ribbon cable. Then, sandwich the touchscreen between the Screen Mounts using four more M2.5 x 20mm screws (**Figure 8.4G**).
14. Attach the leaf switches associated with each stepper motor to the Syringe Stabilizer on their respective sides using M2 x 12mm screws (**Figure 8.4H**).
15. To stand the PDSP vertically, insert one M5 x 8mm screw into the highest through hole on each of the four Right-Angle Brackets. Attach the compact end-feeder into each of these screws and slide them into the bottom of the T-profile rails, closest to the cell homogenizer holder. Do this for the four Right-Angle Brackets for a total of one each on the left and right, and two in the rear (**Figure 8.3I**).
16. Stand the PDSP up over the breadboard and secure the Right-Angle Brackets onto the base of choice using the $\frac{1}{4}$ " x $\frac{3}{4}$ " long socket screws.

17. Attach the Pi Base on the rear of the PSDP such that it is flush against the top of the T-profile rails (**Figure 8.4I**).
18. Thread the wires neatly using the through holes on the Pi Base before connecting each to the Raspberry Pi according to the Electronic Wiring Schematics (**Figure 8.3**).
19. Slide on and secure the Pi Cover (**Figure 8.4H**) to the rear of the Pi Base using M2.5 x 10mm screws (**Figure 8.4I**).
20. Check that all the screws on the PSDP are appropriately tightened prior to use.

8.6.4 Software setup

To install our software and GUI, we have provided the Raspberry Pi image on our lab website (<http://derisilab.ucsf.edu/>). Download the image onto a MicroSD card. We suggest using the free software Etcher to create bootable SD cards [10]. Once installed, unmount the MicroSD and insert it into the Raspberry Pi. Plug in the Raspberry Pi power cord to a conventional outlet to turn it on. To power the stepper motors, connect a conventional 7.5 V power supply to the barrel jack extension cable attached to the PiPlate Motor Plate. Alternatively, install the latest release of Raspbian for the Raspberry Pi, the MOTORplate drivers (via the PIP Python repository: `sudo pip install pi-plates`), and the source code for the PSDP (Supplementary File 3 available with the original publication) and interface manually. The resolution of the Pi can be set to fit any screen by altering the `/boot/conFigure.txt` file. For the touchscreen used here, add the following lines to the end of the `conFigure.txt` file:

```
max_usb_current=1
hdmi_group=2
hdmi_mode=1
hdmi_mode=87
hdmi_cvt 800 480 60 6 0 0 0
```

8.7 OPERATION INSTRUCTIONS

We have designed a GUI optimized for safe and streamlined operation of the PDSP as an automated cell homogenizer (**Figure 8.5**). The software was written with internal limitations to protect both the sample being lysed and the PDSP components. For the push-pull pumping motion, we strongly encourage users to consider employing our intuitive GUI to operate the PDSP. The complete operation instructions of the PDSP as an automated cell homogenizer is described below and can be seen as a video clip in **Supplementary File 2**, which is available with the original publication.

8.7.1 Operation instructions for the provided GUI

1. Turn on the Raspberry Pi by plugging it in. Once the Desktop is loaded, the terminal application will launch and open the provided GUI in full-screen automatically (**Figure 8.5**).
2. Once the GUI has launched, press the “HOME” button to bring both Plunger Movers to their starting position against the Syringe Stabilizer. Pay attention to the popup alerts and ensure that the PDSP is empty before homing.
3. Load a minimum of 1 mL of sample into the cell homogenizer as usual. Make sure the sample is in only one syringe before loading it onto the PDSP.
4. Using the left and right arrow keys under “Volume Control” set the volume of sample in the loaded syringe. Once again, follow the popup alert instructions and ensure the cell homogenizer and syringes have not yet been placed onto the PDSP.
5. Place a filled ice bucket underneath the Cell Homogenizer Holder and insert the cell homogenizer onto the device. Ensure that each syringe is held appropriately by both the Syringe Stabilizer and the left or right Plunger Movers.

6. Using the left and right arrow keys under “Cycle Control,” set the desired number of push-pull cycles. If the cycle number is not explicitly set, the software will default to 20 cycles.
7. Press “*START*.” The number of elapsed cycles along with two timers, a countdown timer based on the estimated duration and a time elapsed counter, should appear.
8. When in doubt, there is an emergency “*STOP*” button that will immediately stop and reset all motors. To reinitialize, remove the cell homogenizer and syringes from the PDSP. Then, return to step 2 outlined here.

8.7.2 Customization suggestions for operation

While we use the PDSP to automate *P. falciparum* lysis using an Isobiotech Cell Homogenizer, the PDSP is adaptable to other tasks. As previously described, the PDSP is programmed on and executed from a Raspberry Pi running the Raspbian operating system. Our software to control the PDSP is written in Python using commands from the PiPlates MOTORPlate Users Guide documentation [11]. Our object-oriented GUI is also designed in Python using Tkinter.

To customize the PDSP for a variety of other applications, users can reference the script that we have written as well as the extensive documentation provided by PiPlate. The PiPlate MOTORplate is highly customizable, offering a wide variety of options for stepper movement in terms of stepper size, speed, and acceleration or deceleration. Furthermore, the PiPlates MOTOR library is highly compatible with the GPIO control library, allowing flexibility in the control of the stepper motors. While python scripts can be executed from the Raspberry Pi command line using a keyboard, users can also create a specialized GUI for their own purposes.

The PDSP presented here is already compatible with a number of different pumping systems. These include, but are not limited to, continuous infusion systems, dual injection systems, and inverse linear constant flow systems. While this PDSP is relatively simple, we can add complexity by attaching up to eight PiPlate MOTORplates powered by one Raspberry Pi, allowing for control of up to sixteen stepper motors simultaneously.

8.8 VALIDATION AND CHARACTERIZATION

8.8.1 Characterization of the PDSP as a syringe pump

To characterize the PDSP, we tested the minimum and maximum flow rates provided by the stepper motors on a 3 mL syringe. By timing the flow of liquid, we were able to measure the consistency and the dynamic range of flow rates for the PDSP (**Figure 8.6A**). We compared these values with the theoretical minimum and maximum flow rates that were calculated as described below:

Let the syringe cross-sectional area have units of mm² and be defined as:

$$A = \pi(\text{syringe diameter}/2)^2$$

Let the linear distance conversion factor (C) have units of mm/degree and be defined as:

$$C = \text{Thread distance}/360^\circ$$

For a T8 threaded rail:

$$C = 8\text{mm}/360^\circ = 1\text{mm}/45^\circ$$

Thus, we can calculate the theoretical flow rate as follows:

$$\text{Flow Rate } (\mu\text{L}/\text{sec}) = A(\text{mm}^2) \times \text{stepsize } (^\circ/\text{step}) \times \text{steprate } (\text{step}/\text{sec}) \times C(\text{mm}/^\circ)$$

The measured flow rates were consistent with the theoretical flow rates calculated using the manufactured syringe diameters and the stepper motor step sizes. This gives us confidence in the PDSP as an appropriate alternative to commercial syringe pumps. Here we provide the

dynamic range of the PDSP in terms of the theoretical minimum and maximum flow rates of various compatible syringe sizes (**Table 8.3**).

8.8.2 Validation of the PDSP as a cell homogenizer

To demonstrate the ability of the PDSP to replace manual cell lysis we compared the two methods directly. In brief, *Plasmodium falciparum* parasites were harvested as previously published [1] and split into two pools, one for manual lysis and one for lysis using the PDSP. We either passed the lysate back-and-forth through the Isobiotec cell homogenizer for 20 cycles by hand or the PDSP controlled by the GUI interface passed the lysate through. Lysates were then collected and centrifuged at 13,000g for 10min at 4C. The supernatant was aliquoted and flash frozen in liquid nitrogen.

One aliquot was taken for each lysis method and used in an *in vitro* translation reaction as previously published with a few modifications [1]. Each 10 μ L reaction consisted of 7 μ L of lysate, 10mM amino acid mixture, 20mM HEPES/KOH pH 8.0, 75mM KOAc, a range of 1–5mM Mg(OAc)₂, 2mM DTT, 0.5mM ATP, 0.1mM GTP, 20mM creatine phosphate, 0.2 μ g/ μ L creatine kinase, and 0.5pmol of Nanoluciferase reporter RNA. Reactions were done in triplicate, incubated at 37C for 45min, and stopped with the addition of 10 μ M cycloheximide. The Nanoluciferase reporter RNA consists of the 130 base pairs directly 5' of PFE_1248300 followed by the Nanoluciferase coding sequence [12] in the 3' UTR of HRP. All RNA was generated off of plasmid digested with PVUII and APALI using T7 transcription (**Figure 8.6B**). After transcription, the DNA template was digested with Turbo DNase (Thermo Fisher) and the RNA was purified using RNA Clean and Concentrate-25 kit (Zymo). Nanoluciferase reactions were performed using Promega's Nano-Glo Luciferase Assay System. In brief, 10 μ L of 1:50 Nanoluciferase substrate:Nanoluciferase buffer was added to each reaction and incubated at room temperature for a minimum of 5min before reading with a 6sec integration on a Promega GloMax-Multi microplate reader.

Our results showed that lysates generated by hand and by the PDSP performed similarly, indicating the PDSP can be used to replace manual lysis for generating in vitro translation extracts (**Figure 8.6C**).

8.8.3 Conclusion and device overview

We have constructed a programmable syringe pump that can be used for biological life science applications. Not only is the PDSP more affordable than commercially available options, but it is also modular and programmable, allowing the user to customize the device for specific tasks or experiments. Here, we have demonstrated that the PDSP can be used to automate and standardize the time-consuming task of *P. falciparum* lysis. Overall, the PDSP can be used as an affordable and customizable alternative to traditional syringe pumps to automate any number of routine laboratory practices.

8.9 ACKNOWLEDGEMENTS AND FUNDING

We would like to acknowledge Wesley Wu for his contributions to the GUI design and Eric Lam for assisting with the 3D printing. This project was funded by the Chan Zuckerberg Biohub, San Francisco, California, United States of America. Valentina Garcia is funded by the University of California TETRAD graduate program and Jamin Liu is funded by the UC Berkeley-UCSF Graduate Program in Bioengineering

8.10 FIGURES

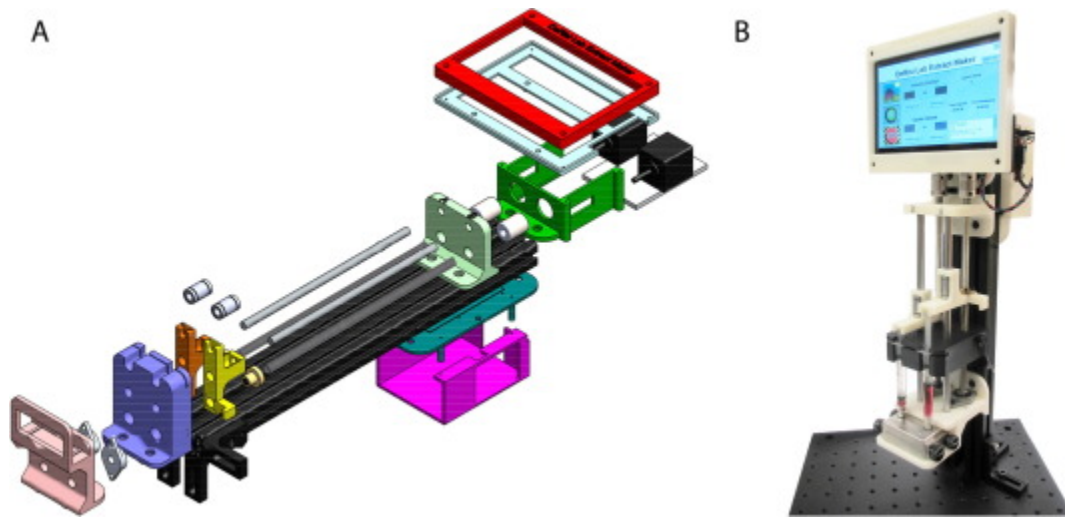


Figure 8.1 Assembled PDSP

A) A 3D rendering of the PDSP with all components visible. **B)** A photograph of the completed PDSP standing vertically.

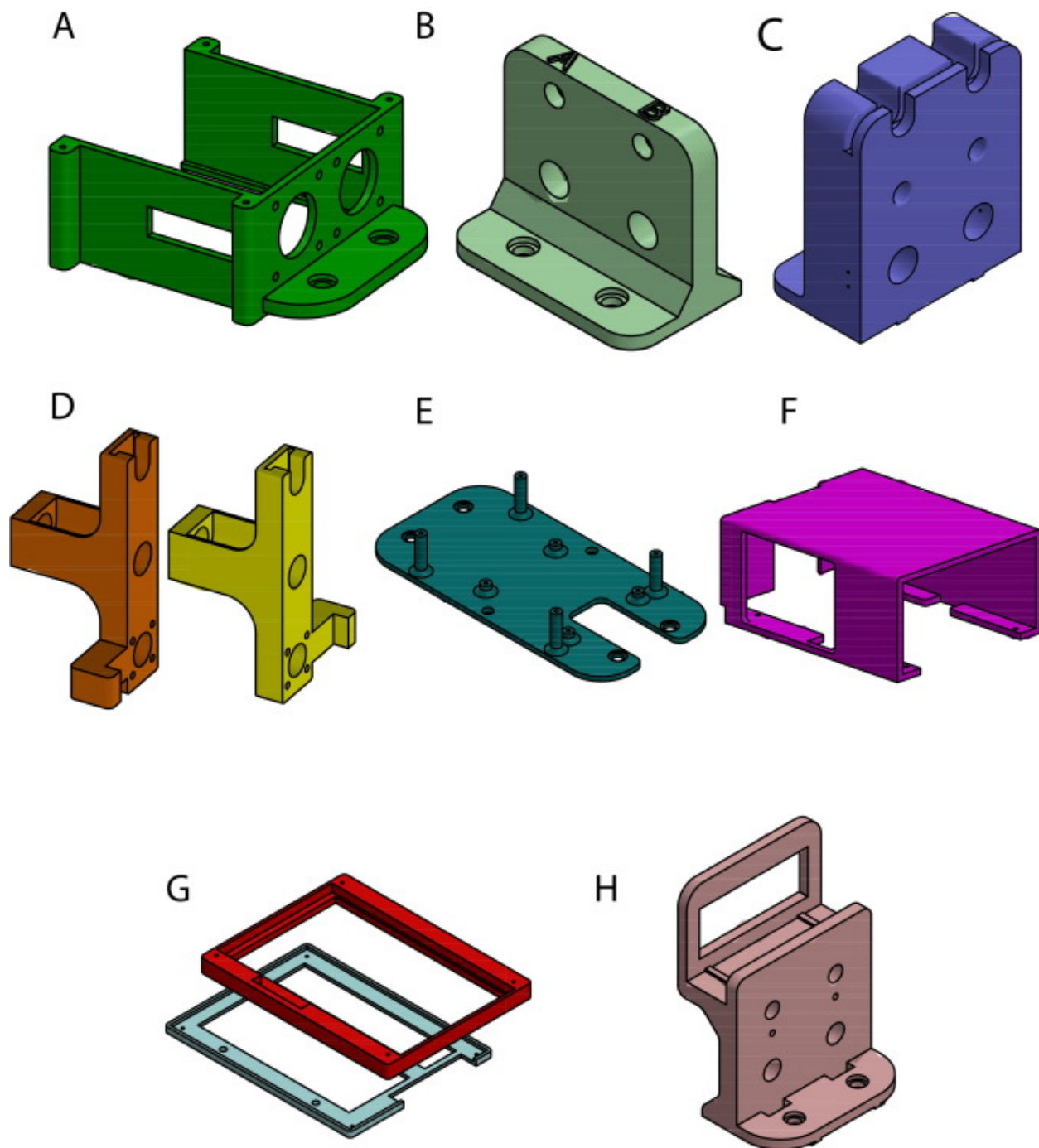


Figure 8.2 3D render of all 3D printed PDSP components

Models of all 3D printed PDSP components: **A)** Motor Base, **B)** Toe Hold, **C)** Syringe Stabilizer, **D)** Left and right Plunger Movers, **E)** Pi Base, **F)** Pi Cover, **G)** Top and bottom Screen Mounts, and **H)** Cell Homogenizer Holder.

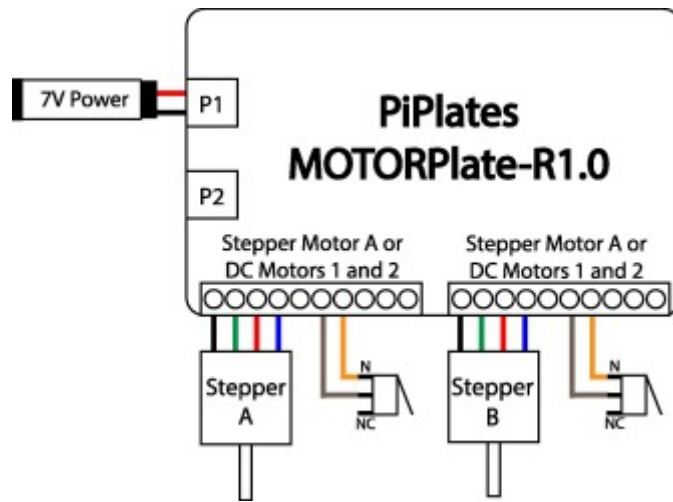


Figure 8.3 Electrical wiring schematic of the PDSP

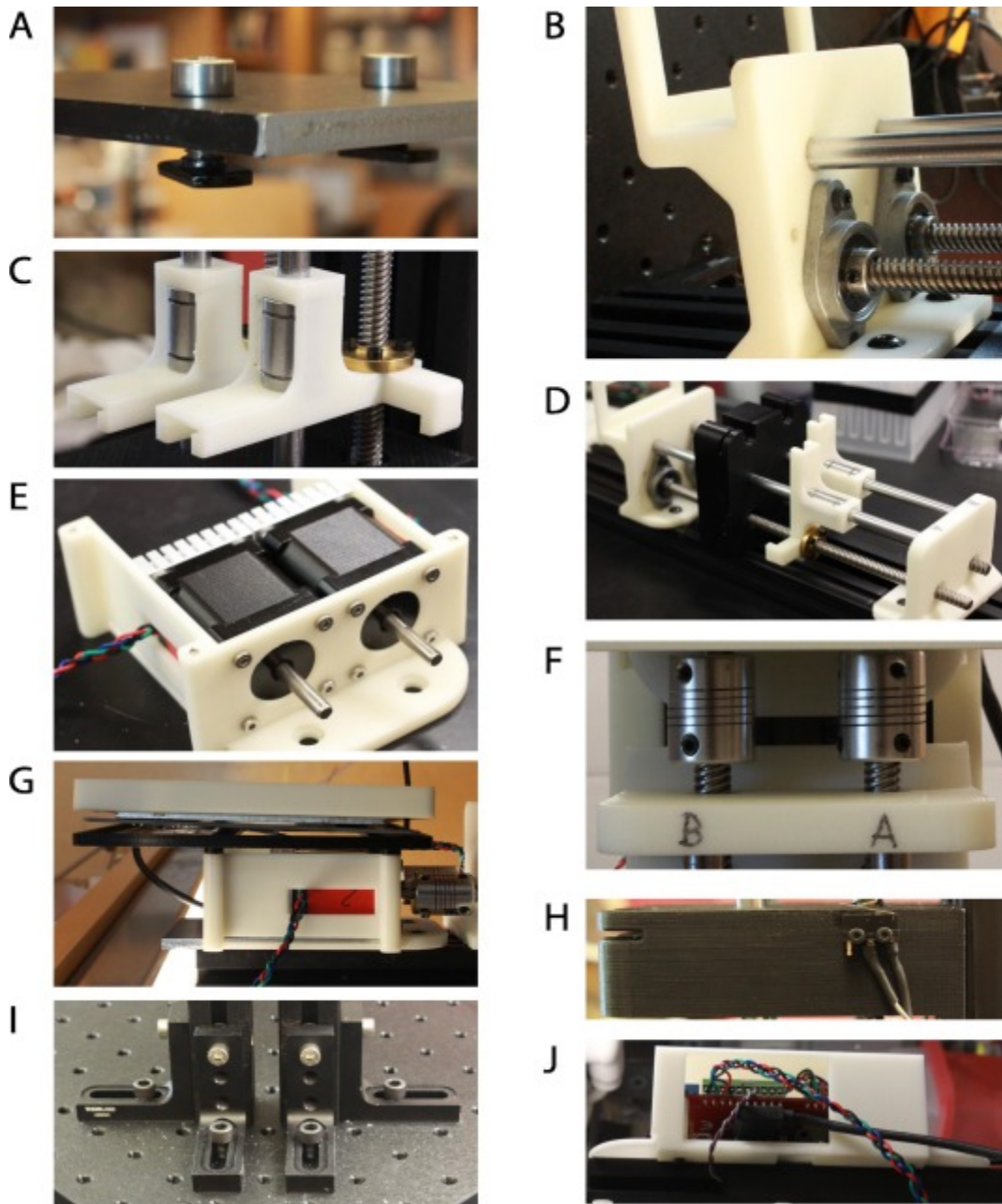


Figure 8.4 Key PDSP assembly steps as described in 8.6.3 Hardware assembly

A) Example of the compact end-feed fastener assembly. **B)** The pillow block bearings, the threaded rods, and the smooth rods attached to the Extract Maker Holder. **C)** the copper nut and linear motion shaft assembled with the Syringe Movers on the smooth and threaded rods. **D)** The Toe Hold, the Syringe Holder, and the Cell Homogenizer Holder assembled on the extruded rails. **E)** The two stepper motors assembled with the Motor Base. **F)** The couplers attached to each motor and the threaded rods. **G)** The Screen Mount Top and Bottom sandwiching the touchscreen and assembled on the Motor Base. **H)** The limit switches attached to the Syringe Stabilizer component. **I)** The Right-Angle Brackets positioned on the aluminum breadboard to stand the PDSP vertically. **J)** The Raspberry Pi and PiPlates Motor Plate attached to the Pi base.

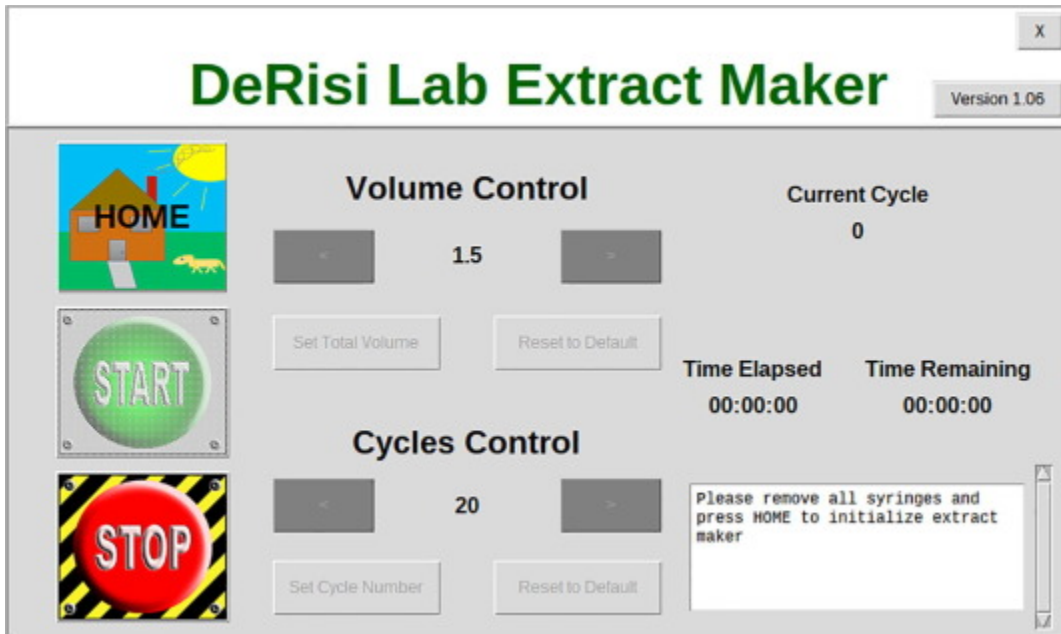


Figure 8.5 PDSP GUI

Interface designed in Python using Tkinter for use of the PDSP with the cell homogenizer.

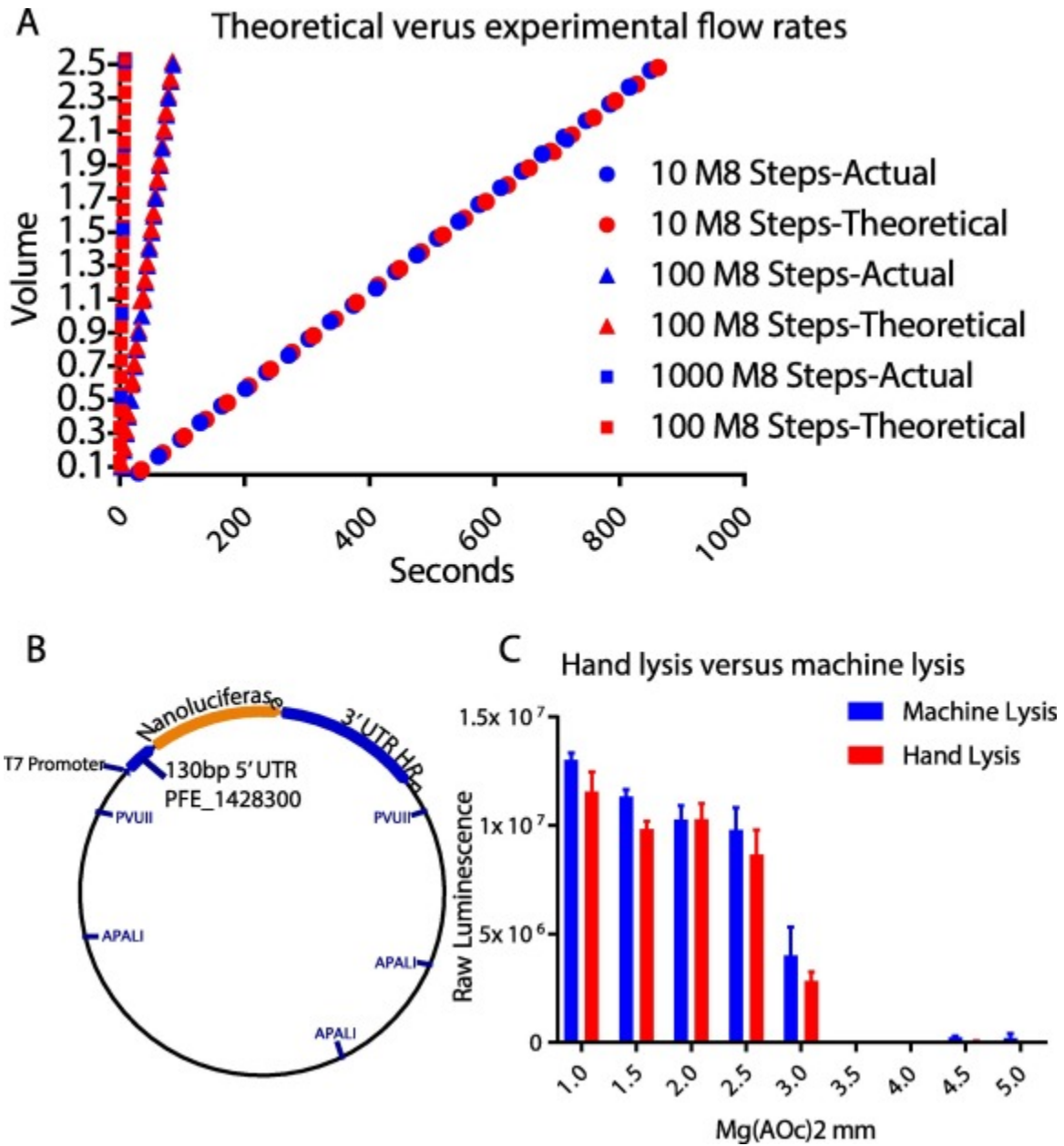


Figure 8.6 PDSP characterization for lysate generation

A) Dynamic Range of Flow Rate of PDSP with 3 mL Syringes. B) Plasmid construct for validation of translational activity in the lysates. C) Translational activity, measured by total luciferase signal, of the lysates generated by hand or the PDSP.

Table 8.1 Design files for all 3D printed parts

Design file name	File type	Open source license	Location of the file
Motor Base (Figure ChapterX.2A)	STL	CC-BY-SA 4.0	https://3dprint.nih.gov/discover/3dpx-008783
Toe hold (Figure ChapterX.2B)	STL	CC-BY-SA 4.0	https://3dprint.nih.gov/discover/3dpx-008783
Syringe Stabilizer (Figure ChapterX.2C)	STL	CC-BY-SA 4.0	https://3dprint.nih.gov/discover/3dpx-008783
Plunger Mover-Left (Figure ChapterX.2D)	STL	CC-BY-SA 4.0	https://3dprint.nih.gov/discover/3dpx-008783
Plunger Mover-Right (Figure ChapterX.2D)	STL	CC-BY-SA 4.0	https://3dprint.nih.gov/discover/3dpx-008783
Pi Base (Figure ChapterX.2E)	STL	CC-BY-SA 4.0	https://3dprint.nih.gov/discover/3dpx-008783
Pi cover (Figure ChapterX.2G)	STL	CC-BY-SA 4.0	https://3dprint.nih.gov/discover/3dpx-008783
Screen Mount-Bottom (Figure ChapterX.2G)	STL	CC-BY-SA 4.0	https://3dprint.nih.gov/discover/3dpx-008783
Screen Mount-Top (Figure ChapterX.2G)	STL	CC-BY-SA 4.0	https://3dprint.nih.gov/discover/3dpx-008783
Cell Homogenizer Holder (Figure ChapterX.2H)	STL	CC-BY-SA 4.0	https://3dprint.nih.gov/discover/3dpx-008783
Right-Angle Bracket	STL	CC-BY-SA 4.0	https://3dprint.nih.gov/discover/3dpx-008783
Aluminum Heat Sink Plate	STL	CC-BY-SA 4.0	Supplementary attachment

Table 8.2 List of all purchased items with number required and cost

Designator	Component	#	Cost per unit (USD)	Total cost (USD)	Source of materials	Material type
<i>Hardware</i>						
M2 × 0.4 mm	Thread × 12 mm Long Socket Head Screw	4	\$0.12	\$0.47	McMaster-Carr	Steel
M2.5 × 0.45 mm	Thread × 20 mm Long Socket Head Screw	8	\$0.28	\$2.24	McMaster-Carr	Steel
M2.5 × 0.45 mm	Thread × 10 mm Long Socket Head Screw	16	\$0.15	\$2.45	McMaster-Carr	Steel
M2.5 × 0.45 mm	Thread × 5 mm Phillips Flat Head Screw	4	\$0.03	\$0.13	McMaster-Carr	Steel
M3 × 0.5 mm	Thread × 5 mm Long Socket Head Screw	8	\$0.08	\$0.66	McMaster-Carr	Steel
M3 × 0.5 mm	Thread × 20 mm Long Socket Head Screw	2	\$0.10	\$0.20	McMaster-Carr	Steel
M5 × 0.8 mm	Thread × 8 mm Long Socket Head Screw	12	\$0.09	\$1.08	McMaster-Carr	Steel
M5 × 0.8 mm	Thread × 15 mm Long Socket Head Screw	6	\$0.15	\$0.88	McMaster-Carr	Steel
1/4"-20	Thread × 3/4" Long Socket Head Screw	4	\$0.15	\$0.60	McMaster-Carr	Steel
Compact End-Feed Framing	Fastener, M5, for 20 mm T-Slotted	18	\$0.50	\$8.96	McMaster-Carr	Steel
1.00" × 1.00"	Smooth Surface T-Slotted Profile	2	\$7.20	\$14.40	80/20 Inc.	Aluminum
Aluminum Sheet	– 71 × 90 × 5 mm	1	\$21.42	\$21.42	OnlineMetals	Aluminum
T8 Bearings	Lead Screw + Coupler + Pillow Block Lead Screw Set (300 mm)	2	\$12.52	\$25.04	Amazon	Steel
Linear Motion	8 mm Shaft × 330 mm Length	2	\$15.18	\$30.36	Amazon	Steel
Linear Bearing	Ball Bushing	2	\$1.21	\$2.43	Amazon	Steel

Designator	Component	#	Cost per unit (USD)	Total cost (USD)	Source of materials	Material type
<i>Electronics</i>						
Raspberry Pi 3 – Model B – ARMv8 with 1G RAM	3055	1	\$35.00	\$35.00	Adafruit	Other
5 V 2.4A Switching Power Supply with 20AWG MicroUSB Cable	1995	1	\$7.50	\$7.50	Adafruit	Other
16 GB microSDHC Class 10 Flash Memory Card	415,141	1	\$7.99	\$7.99	Microcenter	Other
7" Touch Screen for Raspberry Pi	B01ID5BQTC	1	\$42.88	\$42.88	Amazon	Other
Right HDMI Male to Left HDMI Male Cable	B06XT2JS1G	1	\$7.99	\$7.99	Amazon	Other
MOTORplate	MOTORR1	1	\$39.99	\$39.99	Pi Plates	Other
2.1 mm female/male barrel jack extension cable	327	1	\$2.95	\$2.95	Adafruit	Other
7.5 Volt 2.4 Amp Power Adapter	B006QYXFRO	1	\$21.99	\$21.99	Amazon	Other
Nema 14 Bipolar 1.8 deg 23Ncm 0.5A 7.5 V 35x35x42mm	14HS17-0504S	2	\$14.27	\$28.54	StepperOnline	Other
Omron Snap Action Switch	COM-00098	2	\$1.95	\$3.90	Sparkfun	Other
Heatsink 35 mm x 35 mm x 10 mm	B00MJVXB9K	2	\$1.39	\$2.77	Amazon	Aluminum
Heat Sink Thermal Tape 80 mm x 80 mm	1468	1	\$4.50	\$4.50	Adafruit	Other
Wire Assortment	38,687	1	\$9.99	\$9.99	Microcenter	Other
Heat Shrink Tubing Assortment	544,023	1	\$2.99	\$2.99	Microcenter	Other
Solder	860,585	1	\$16.99	\$16.99	Microcenter	Other
<i>Optional</i>						
Aluminum Breadboard 12" x 12" x 1/2", 1/4"-20 Taps	MB12	1	\$151.00	\$151.00	Thorlabs	Aluminum
Slim Right-Angle Bracket with Counterbored Slot & 1/4"-20 Tapped Holes	AB90C	4	\$26.25	\$105.00	Thorlabs	Aluminum

Table 8.3 Minimum and maximum theoretical flow rates by the PDSP based on syringe size

Syringe Size	Diameter (mm)	Minimum Flow Rate ($\mu\text{L/s}$)	10 Steps/s ($\mu\text{L/s}$)	100 Steps/s ($\mu\text{L/s}$)	Maximum Flow rate (mL/S)
<i>M8 Step size</i>					
1 mL	4.78	0.1	0.9	9.0	0.090
3 mL	8.66	0.3	2.9	29.5	0.295
5 mL	12.06	0.6	5.7	57.1	0.571
10 mL	14.5	0.8	8.3	82.6	0.826
20 mL	19.13	1.4	14.4	143.7	1.437
30 mL	21.7	1.8	18.5	184.9	1.849
50/60 mL	26.7	2.8	28.0	280.0	2.800
 <i>Full Step Size</i>					
1 mL	4.78	0.7	7.2	71.8	0.718
3 mL	8.66	2.4	23.6	235.6	2.356
5 mL	12.06	4.6	45.7	456.9	4.569
10 mL	14.5	6.6	66.1	660.5	6.605
20 mL	19.13	11.5	115.0	1149.7	11.497
30 mL	21.7	14.8	147.9	1479.3	14.793
50/60 mL	26.7	22.4	224.0	2239.6	22.396

8.12 REFERENCES

1. V. Ahyong et al, Identification of Plasmodium falciparum specific translation inhibitors from the MMV Malaria Box using a high throughput in vitro translation screen, *Malar. J.* 15 (2016) 173.
2. S. Bhaskaran et al, Breaking Caenorhabditis elegans the easy way using the Balch homogenizer: an old tool for a new application, *Anal. Biochem.* 413 (2011) 123–132.
3. D. Sarnataro et al, PrP(C) association with lipid rafts in the early secretory pathway stabilizes its cellular conformation, *Mol. Biol. Cell* 15 (2004) 4031– 4042.
4. SyringePump.com – Push-Pull Continuous Infusion, Dual Pump System. Available at: <https://www.syringepump.com/special.php> (accessed 14th March 2018).
5. J.R. Lake, K.C. Heyde, W.C. Ruder, Low-cost feedback-controlled syringe pressure pumps for microfluidics applications, *PLoS One* 12 (2017) e0175089.
6. B. Wijnen, E.J. Hunt, G.C. Anzalone, J.M. Pearce, Open-source syringe pump library, *PLoS One* 9 (2014) e107216.
7. M.C. Carvalho, R.H. Murray, Osmar, the open-source microsyringe autosampler, *HardwareX* 3 (2018) 10–38.
8. K. Pusch, T.J. Hinton, A.W. Feinberg, Large volume syringe pump extruder for desktop 3D printers, *HardwareX* 3 (2018) 49–61.
9. J.Z. Milanovic, P. Milanovic, R. Kragic, M. Kostic, 'Do-It-Yourself' reliable pH-stat device by using open-source software, inexpensive hardware and available laboratory equipment, *PLoS One* 13 (2018) e0193744.
10. Installing Operating System Images – Raspberry Pi Documentation. Available at: <https://www.raspberrypi.org/documentation/installation/installingimages/> (accessed 14th March 2018).
11. Code, Pi-Plates, 2015.

12. M.P. Hall et al, Engineered Luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate, ACS Chem. Biol. 7 (2012) 1848– 1857.

9 APPENDIX I: DATA AVAILABILITY

Below is a reference for all sequencing data and supplementary files and their availabilities in public repositories. Resources are ordered in accordance with their relevant chapters.

CHAPTER 2

Description	Source
Raw amplicon sequencing of sgRNA of sorted and unsorted A549 cells following infection with LCMV reporter virus.	NCBI BioProject Accession ID PRJNA806912
Hit list supplementary file	

CHAPTER 5

Description	Source
Raw Artic_v3 sequencing .fastq files from a secondary nasal swab collected following a positive Abbot BinaxNOW rapid antigen test, collected with Unidos en Salud from the Mission District	CZID.org/public Project names: 20210119_MissionBinax_SpikeOnly Mission Bridges Study January 10th-January 30th
Consensus Genome assembled from Artic_v3 amplicon sequencing as a .fasta file.	GISAID Repository Virus names: hCoV-19/USA/CA-UCSF-JD1/2021 to hCoV-19/USA/CA-UCSF-JD929/2021

CHAPTER 6

Description	Source
Consensus Genome assembled from Artic_v3 amplicon sequencing as a .fasta file.	GISAID Repository Virus names can be found in Table 6.S1

CHAPTER 7

Description	Source
Raw Artic_v3 or Midnight sequencing .fastq files from a nasal swab by either Unidos en Salud from the Mission District or by the hospital clinical laboratory.	NCBI BioProject Accession ID PRJNA790937
Concensus Genome assembled from Artic_v3 or Midnight amplicon sequencing as a .fasta file.	GISAID Accession: EPI_ISL_2508365 EPI_ISL_2508366

CHAPTER 8

Description	Source
Supplementary File 1 (.stl) Aluminum heat sink dimensions	https://www.hardware-x.com/cms/10.1016/j.ohx.2018.e00027/attachment/e47df766-3242-40ea-a3a9-3acbc45b7050/mmc1.zip
Video Supplementary File 2 (.mp4) Time-lapse of PDSP assembly	https://www.hardware-x.com/cms/10.1016/j.ohx.2018.e00027/attachment/e4bc5bba-773f-418f-a51a-7bb8f0055ac1/mmc2.mp4
Supplementary File 3 (various) Source code for the PDSP	https://github.com/liujamin/pdsp_extractmaker

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:

Chieh Ming Liu

AA7D54D885B14D2...

Author Signature

6/15/2022

Date