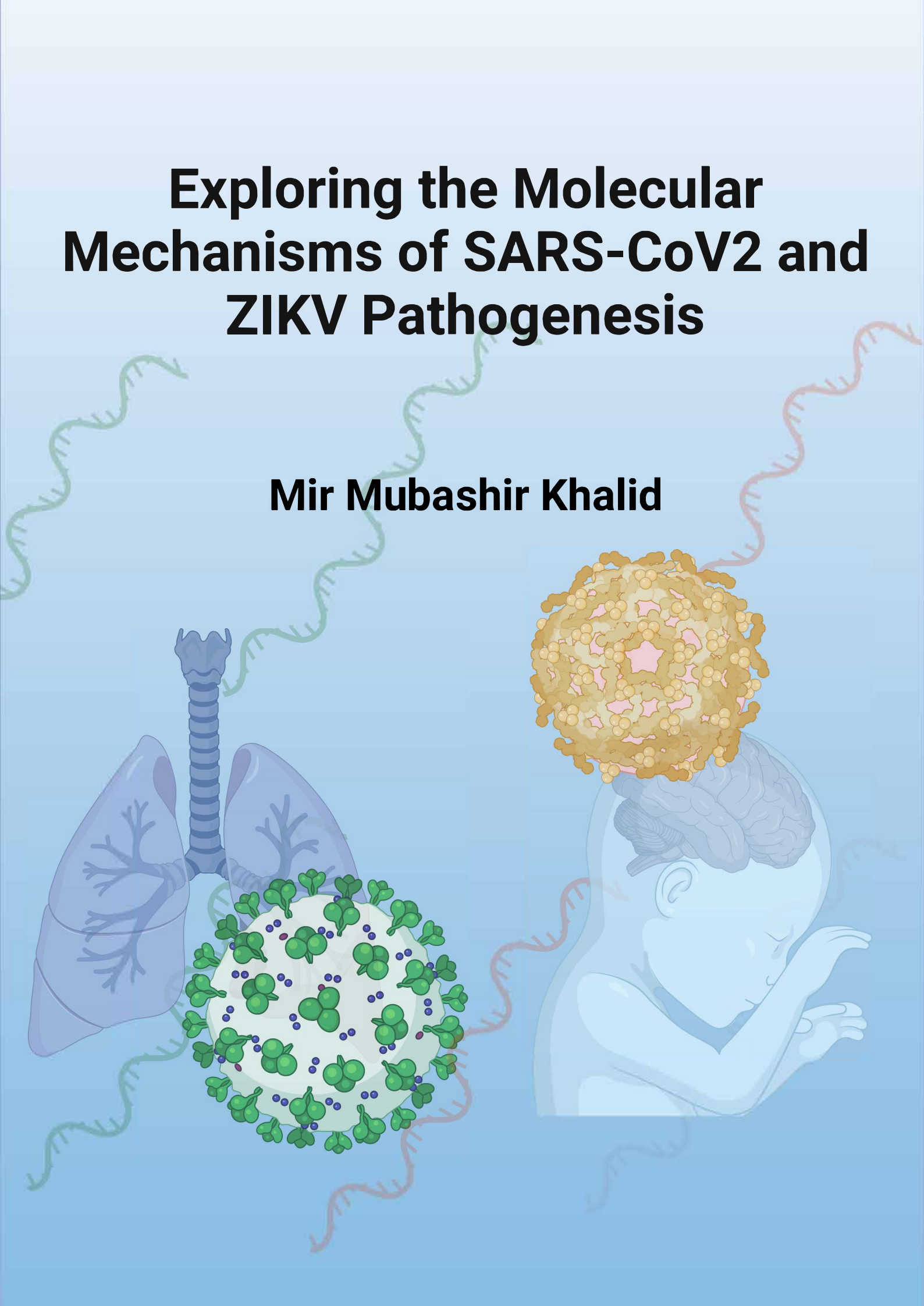
Exploring the Molecular Mechanisms of SARS-CoV2 and ZIKV Pathogenesis

Mir Mubashir Khalid



Exploring the molecular mechanisms of SARS-CoV2 and ZIKV pathogenesis

Mir Mubashir Khalid

ISBN
978-94-6361-624-9

The research in this thesis was performed at the Gladstone Institute of Virology, Gladstone Institutes (UCSF), San Francisco, CA, The United States. The candidate is enrolled Medisch Genetisch Centrum Zuid-West Netherland (MGC) PhD program in the Erasmus University Medical Center, Rotterdam, The Netherlands.

Cover illustrations
ssRNA virus SARS-CoV2 infects lungs, and ZIKV infects fetal brains thus causing pathogenesis.

Cover design
Mir Mubashir Khalid. (Created with BioRender.com)

Printed by
Optima Grafische Communicatie, Rotterdam / www.ogc.nl

Exploring the Molecular Mechanisms of SARS-CoV2 and ZIKV Pathogenesis

Onderzoek naar de moleculaire mechanismen van de pathogenese van SARS-CoV2 en ZIKV

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on

Tuesday, January 11th, 2022 at 15:30 hrs
by

Mir Mubashir Khalid
Born in Gazipur, Bangladesh

Erasmus University Rotterdam

Doctoral Committee:

Promotor: **Prof.dr. C.P. Verrijzer**

Other members:

Prof.dr. E. Verdin
Dr. B. Haagmans
Dr. N. Roan

Copromotors: **Prof.dr. T. Mahmoudi**
Prof.dr. M. Ott

To my family, wife, friends
and all science enthusiasts

Table of contents

Figures and Table

Abbreviations

ACE2= Angiotensin-converting enzyme 2

ARDS= Acute Respiratory Distress Syndrome

AXL= AXL Receptor Tyrosine Kinase

CDC= Ce Centers for Disease Control and Prevention

CHIKV= Chikungunya Virus

CoV= Coronavirus

COVID-19= Coronavirus Disease 2019

DC-SIGN= dendritic cell-specific ICAM-grabbing non-integrin DENV=Dengue Virus

DNA= Deoxyribonucleic Acid

EJC= Exon-junction complex

FDA= Food and Drug Administration

FREM2= FRAS1 Related Extracellular Matrix 2

GBS= Guillain Barre syndrome

GISAID= Global Initiative on Sharing Avian Influenza Data

GWAS= genome-wide association study

HCV= Hepatitis C Virus

HSV-1= Herpes simplex virus-1

IFN= Interferon

IL-1β = Interleukin 1 beta

IL17RA= Interleukin 17 Receptor A

IL17RB= Interleukin 17 Receptor B

IL17RC= Interleukin 17 Receptor C

IL17RD= Interleukin 17 Receptor D

IL17RE= Interleukin 17 Receptor E

IL17A= Interleukin 17 A

IL17A= Interleukin 17 F

MERS= Middle East Respiratory Syndrome

MHC-I= Major histocompatibility complex class I

miRNA= microRNA

NMD= Non-sense Mediated RNA Decay

NPC= Human neural progenitor cells

NSC= Neural stem and progenitor cells

Orf8= Open reading frame-8 protein

PTC= Premature termination codons

RBD= Receptor Binding Domain

RNA= Ribonucleic Acid

RSV= Respiratory Syncytial Virus

SARS=Severe Acute Respiratory Syndrome

sIL17RA= soluble IL17RA

SIV=Simian Immunodeficiency Virus

ssRNA= single stranded RNA

TGF-β = Transforming growth factor beta

Th17= T helper 17 cell

TIM/TAM= T cell immunoglobulin domain and mucin domain/ Tyro3, Axl, and Mer

TM= Transmembrane domain

TMPRSS2= Transmembrane protease, serine 2

UPF1= Up-frameshift suppressor 1 homolog

WNV= West Nile Virus

YFV=Yellow Fever Virus

ZIKV=Zika Virus

Chapter 1

General Introduction

General Introduction

Epidemiology of RNA viruses

Viral pandemics and epidemics are not new phenomena in human history. Prior to modern medicine, humans have experienced continuous threats of plagues and outbreaks. From the deadliest black plague to the current SARS-CoV2 pandemic, millions of lives have been devastated by bacteria, viruses, and parasites. Viral infection or outbreak has been traced back to ancient Egypt (~1157 BCE) (1). At those times, evil spirits, supernatural causes, and punishment by God were blamed for such outbreaks (2, 3). Lack of understanding of how infectious disease spreads and how to apply successful preventive measures are primary reasons for recurring outbreaks. One devastating viral pandemic in history was smallpox, which took more than ~300 million lives (4). After a successful vaccination campaign against smallpox that eradicated the virus in 1980, the general population started to believe that outbreaks and pandemics were preventable by modern medicine. In recent years, rapid urbanization and advances in transportation caused drastic changes to the environment and ecology. In addition, deforestation destabilizes animal habitats bringing human and wild animals in close contact. Human and wild animal interaction increases the chances of a virus, which present in animals, spreading to humans. This event is called the spillover event (5, 6). Viral transmission from an animal reservoir to a human is called zoonotic transmission of the virus. Arthropod vectors such as mosquitoes or ticks can also spread viruses to humans, thus causing vector-borne virus transmission. With faster transportation and global trade advancement, an infected person or animal can be transported to different parts of the world, thus can spread a disease globally. The threat of epidemics and pandemics is still at large even these days.

Over the past 20 years, two major viral outbreaks were caused by Coronaviruses (CoV) and Zika virus (ZIKV). Three Coronviruses that caused outbreaks were Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), Middle East Respiratory Syndrome Coronavirus (MERS-CoV), and Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) (7-9). Coronaviruses are frequently found in bats(10). In December 2019, the first case of Coronavirus Disease 2019 (COVID-19) was reported in Wuhan, China (11). SARS-CoV2 is the causal agent of COVID-19. It is suspected that the SARS-CoV2 spread from bats to humans via an intermediate animal host (12). After zoonotic transmission of SARS-CoV2, it quickly spreads globally and is continuing the pandemic. A vector-borne virus, ZIKV spreads mainly via mosquito bites. In March 2015, the ZIKV outbreak originated in South America and quickly spread to North America, the

Pacific, and Southeast Asia and ended in 2016 (13-15). These two viruses are the most recent viral outbreaks that caught global attention. However, there are lots unknown about the biology and features of these two viruses.

SARS-CoV2 and ZIKV have different transmission rates, animal host or vector, and pathology, but both contain an RNA genome. They have a positive-sense single-stranded RNA (+ssRNA) genome and have a similar replication cycle (**Figure 1**). The current COVID-19 pandemic caused a massive death toll and severe complications. The mortality was primarily seen in the older population and adults with pre-existing conditions (e.g., heart disease and diabetes) (16). Younger patients were largely asymptomatic. ZIKV causes severity in fetuses, while infected adults are mainly asymptomatic or mildly ill. Congenital ZIKV infection causes a developmental defect called microcephaly. Microcephaly is characterized by the reduced circumference of head size of the newborn (9). Similarities between these two viruses include that both these viruses encode a polyprotein or multiple polyproteins subjected to proteolytic cleavage and processing and enter cells via receptors (**Figure 1**). Even though both of these RNA viruses have similarities, their transmission and replication are affected differently.

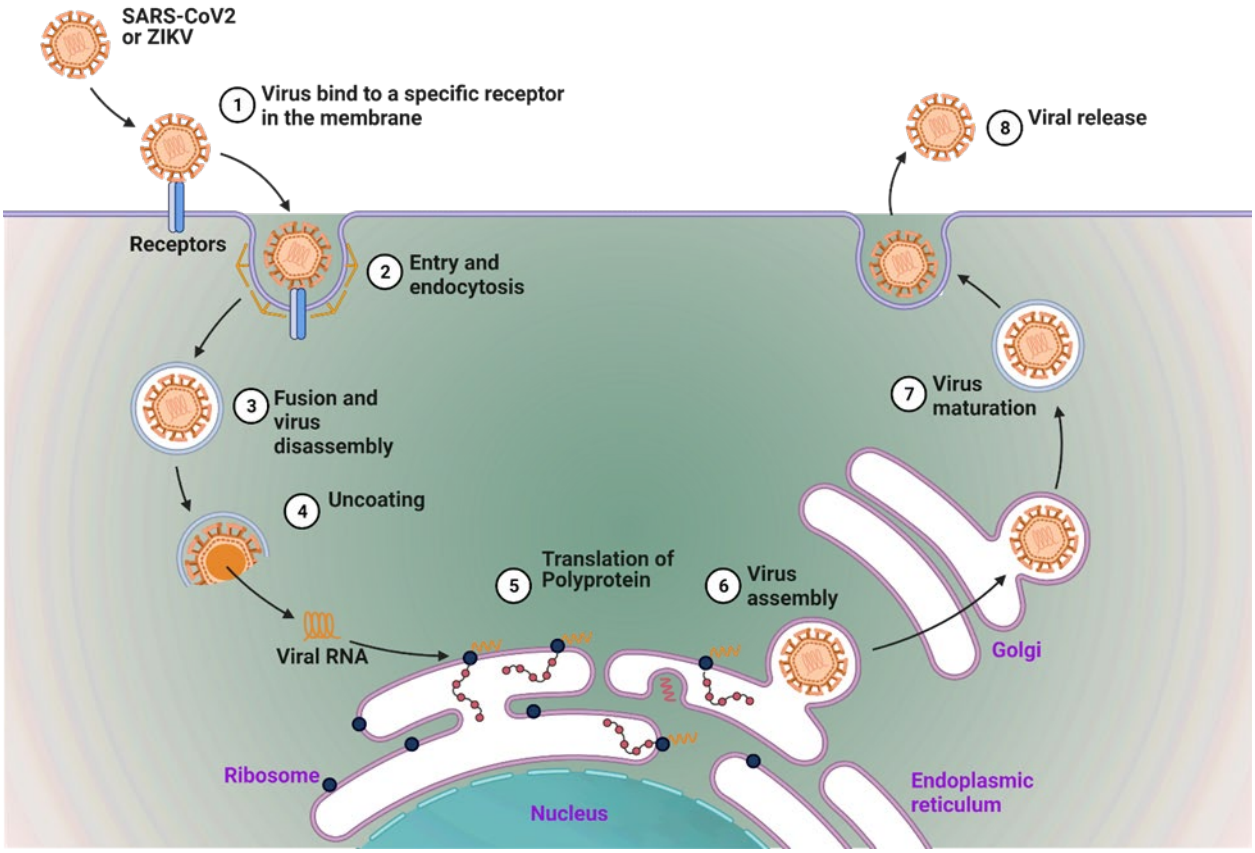


Figure 1: A schematic of SARS-CoV2 and ZIKV replication cycle. Step**1-4**: Both virus targets cells which express their putative receptors. SARS-CoV2 uses angiotensin-converting enzyme 2 (ACE2) and transmembrane protease serine 2 (TMPRSS2). Attachment of S1 subunit of SARS-CoV2 spike protein with ACE2 receptor prepares virus for endocytosis, followed by S2 subunit cleavage by TMPRSS2. This cleavage causes conformational changes, which helps viral entry to the acidified endosome and causes viral RNA release to cytosol. ZIKV uses AXL Receptor Tyrosine Kinase (AXL), Tyro, dendritic cell-specific ICAM-grabbing non-integrin (DC-SIGN), T cell immunoglobulin domain and mucin domain/ Tyro3, Axl, and Mer (TIM/TAM) receptors for entry. ZIKV envelope protein (E) is primarily involved in the viral entry in a similar manner of SARS-CoV2. Unlike SARS-CoV2, ZIKV E protein can change conformation after binding and infection, leading to fusion to the acidic endosome. The low pH environment and conformational changes lead to membrane fusion, ultimately releasing capsid-bound RNA to the cytosol. Step **5-8**: Both viruses follow a cap-dependent translation, leading to a long polyprotein (ZIKV) or multiple polyproteins (SARS-CoV2) subject to co- and post-translationally cleavage by different proteases. Viral replication occurs via non-structural proteins (NSP) with RdRp, an RNA-dependent RNA polymerase. Finally, assembly at the endoplasmic reticulum (ER) followed by maturation leads to mature virion to egress from the cell (17, 18).

This thesis investigates viral factors important for fitness and pathogenesis. First, we investigated the role of two SARS-CoV2 proteins (spike and Open reading frame-8 protein (ORF8)). We identified a new variant (Epsilon) of SARS-CoV2 with unique spike mutations. Then we investigated its transmission, infectivity, and immune evasion. Later we investigated conserved and unique host targets for three coronaviruses. We found the role of SARS-CoV2 ORF8 interaction with Interleukin 17 Receptor A (IL17RA). Secondly, we investigated the role of ZIKV capsid protein on the host RNA surveillance pathway. This thesis dives into the role of these viral proteins to show how viral proteins target specific host factors and pathways to benefit viral replication and cause pathogenesis. Thus, studying the function and cellular interactors of individual proteins at the molecular level will help discover new roles and functions of viral proteins. This knowledge can help to limit, prepare against and prevent viral infection and spread.

1.1 RNA virus outbreak, emergence and spread in 21st century

Previously, a predominant cause of plagues and outbreaks was poor hygiene and lack of medical knowledge. However, medical science and the development of sanitary practices improved hygiene and introduced vaccination and preventive measures to control contagious diseases. This has lead to a rapid decrease in infection-related deaths in developed countries.

However, in the 21st century, we started seeing many viral outbreaks, mostly RNA virus-mediated outbreaks. The current COVID-19 pandemic started in 2019 in Wuhan, China (11). Although Chinese authorities followed a strict lockdown and quarantine protocol, the virus spread rapidly to other parts of the world, underscoring how global connectivity can facilitate the spread of a newly emerged respiratory virus. Additionally, the longer incubation period and asymptomatic infected individuals benefited virus transmission, allowed the virus to disseminate, and thus, it spread to other countries. Other factors that allowed the virus to spread efficiently include reluctance to wear masks, following quarantine guidelines, and spreading false information and confusion through social media (19). At the time of writing this thesis, close to 4 million people died. While more than 180 million got infected worldwide, and continuation of infections are still going (20). This ongoing pandemic took millions of lives and devasted the economy.

The causal agent of COVID-19, SARS-CoV2, belongs to *Coronaviridae* family. The family includes around 46 members of viruses and two subfamilies, the *Orthocoronavirinae* and the *Letovirinae* (21). Based on phylogenetic relationship and genome structure, the *Orthocoronavirinae* subfamily is subdivided into four major genera (*Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus*). The majority of them are known to infect mammals, including seven known to infect humans belonging to the alpha and beta-coronavirus. The SARS-CoV2 belongs to the *Betacoronavirus* genera (22, 23). Human contact with wild animals infected with coronaviruses, such as bats or pangolins, represents a significant source of viruses that may spillover into human populations. Transmission from an animal to a human is the most likely scenario for the source of SARS-CoV-2 (23). Although there has been a recent investigation on the lab leak hypothesis of SARS-CoV2 as opposed to the natural origin, the bat is still considered a natural host where the spillover events to human occurred through another intermediate host (24, 25). The intermediate host for SARS-CoV2 is still under debate; a suspected intermediate host is pangolin (12). SARS-CoV2 is more similar to coronavirus identified in bat than SARS-CoV (26). SARS-CoV2 spreads primarily through droplet transmission, whereas airborne transmission is still debatable (**Figure 2**) (27). In ~1.5 years, millions of people died after SARS-CoV2 infection due to SARS-CoV2 pathogenesis. Upon infection, SARS-CoV2 can lead to severe respiratory conditions such as ARDS (Acute Respiratory Distress Syndrome), leading to death. SARS-CoV2 infection is more severe and likely to be fatal in older adults (**Figure 2**) (28).

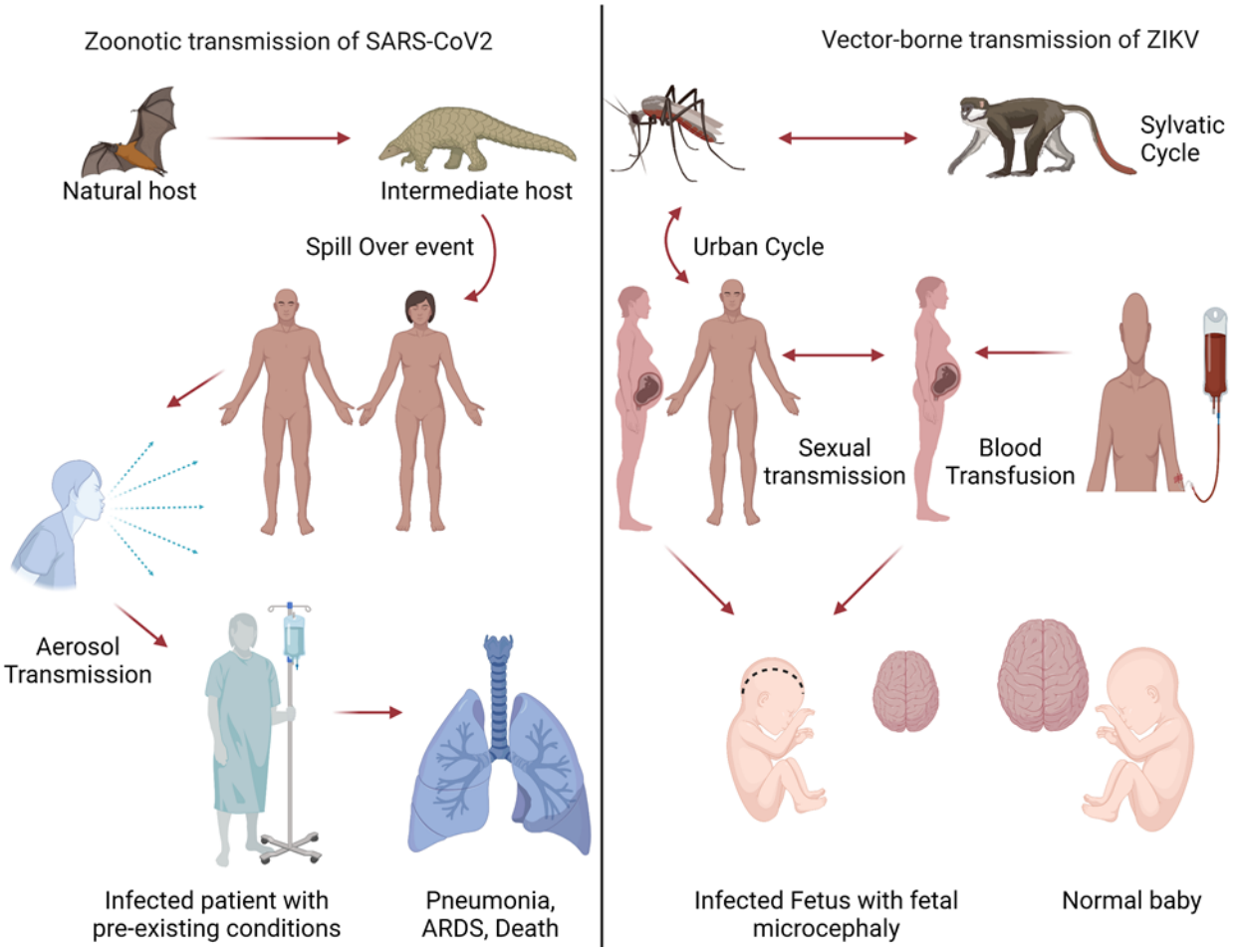


Figure 2: Two different transmission patterns of SARS-CoV2 and ZIKV outbreaks. On the **left**, zoonotic transmission of SARS-CoV2 is described. Bats are natural hosts for Coronaviruses like SARS-CoV2, which can be transmitted to humans through an intermediary host like pangolin (10, 22). After the virus adapts to humans, SARS-CoV2 transmits by respiratory fluids. A person with pre-existing conditions like diabetes or heart disease can develop severe conditions after getting infected with SARS-CoV2. These severe conditions can lead to pneumonia, ARDS, and even death. On the **right**, vector-borne transmission of ZIKV is described. Like other flaviviruses like Dengue Virus(DENV), ZIKV also has a sylvatic cycle which is maintained by the infection cycle between forest-dwelling mosquitoes and non-human primates in the forest. An urban cycle starts when ZIKV infection happens in humans from mosquito bites from forest-dwelling mosquitoes (29). ZIKV can spread to others via mosquito bites, sexual transmission, mother-to-child, and blood transfusion. Congenital ZIKV infection can lead to a reduction of fetal brain size known as fetal microcephaly.

ZIKV is a +ssRNA virus belonging to the flavivirus genus of the *Flaviviridae* family, which includes many arboviruses. Arboviruses are viruses that are transmitted by an arthropod vector, most commonly ticks and mosquitoes. ZIKV is a flavivirus and arbovirus that spreads through mosquito bites. ZIKV has a sylvatic cycle, where the virus circulates in non-human primates and forest-dwelling mosquitoes, and an urban cycle, where the virus spreads in human habitats with mosquitoes that live near humans civilization (**Figure 2**) (29-31). The animal reservoirs of ZIKV make eradication of it especially difficult. A crossover event results in ZIKV circulating in the urban cycle, and the virus starts to adapt in humans (**Figure 2**). When ZIKV bearing mosquitoes are exposed to a new population that has no immunity, viral outbreaks happen. ZIKV is adapted to replicate both humans and mosquitoes. The area where ZIKV outbreaks occur matches accurately with temperate zones. So, ZIKV spread mainly depends on the environment that sustains mosquito growth and human interactions. Beyond transmission by the mosquito, other transmission routes of ZIKV are sexual intercourse, blood transfusion, and mother to child (32, 33). Furthermore, ZIKV can persist in semen for a long time and probably can transmit via sexual intercourse even after the symptomatic period (34). Overall, the multiple routes by which ZIKV can be transmitted to humans and the presence of animal reservoirs makes eradication challenging.

First identified in a Ugandan forest in 1947, ZIKV did not cause significant outbreaks prior to 2007 (35). It generally caused few infections with mild symptoms. The first ZIKV outbreak was reported in Yap Islands in 2007 and later in 2013 in French Polynesia, where ZIKV infection was linked to an increased number of Guillain Barre Syndrome (GBS) (13). GBS is a rare neurological disorder where an autoimmune attack damages the peripheral nervous system. GBS has also been linked to *Campylobacter*, Influenza infection, and other viral infections (36). In early 2015, Brazil experienced a high number of patients with fever caused by ZIKV (14, 15, 37). Concurrently, fetal microcephaly, babies born with a smaller head size than normal, increased in the region. Ultimately, ZIKV infection was linked to this increase in fetal microcephaly (13, 33, 37). In 2015, ZIKV spread to other countries in North America, South America, and Southeast Asia and World Health Organization (WHO) declared ZIKV as a global health concern in February, 2016 (13). Many countries took preventive measures and cautions to stop the epidemic, and in November 2016, WHO declared the end of the ZIKV epidemic. In most adults, ZIKV infection usually results in mild symptoms like fever, joint pain, and headache. Pregnant mothers infected with ZIKV can result in Zika infection in the fetus. ZIKV is able to cross the Blood Brain Barrier(BBB), replicate in the brain, and cause fetal microcephaly (38). ZIKV infection can disrupt the developmental process of brain cells. ZIKV promotes the differentiation of brain cells for its

replicative advantages, which leads to a reduced number of brain cells (39). Fetal microcephaly is characterized by the reduced circumference of fetal head size (**Figure 2**). ZIKV promotes the differentiation of brain cells for its replicative advantages, which leads to a reduced number of brain cells (39). Thus ZIKV infection can cause severe complications in fetuses (**Figure 2**).

1.2 The emergence of new variants in the ongoing SARS-CoV2 pandemic

SARS-CoV-2 is an enveloped virus with a ~30Kb positive-sense ssRNA genome and the causative agent of COVID-19 (26). The 5' end of the SARS-CoV2 genome encodes two long polyproteins that are later proteolytically processed to 16 non-structural proteins (NSP1-16). The 3' end encodes four structural proteins spike (S), envelope (E), membrane (M) and nucleocapsid (N), and nine accessory proteins (ORF3a, ORF3b, ORF6, ORF7a, ORF7b, ORF8, ORF9a, ORF9b, ORF10) (**Figure 3**) (40). Out of 4 structural proteins, spike is essential for viral entry, interacting with host cellular receptors angiotensin-converting enzyme 2 (ACE2) and transmembrane protease serine 2 (TMPRSS2) (26, 41). Spike is one of the most variable regions in the SARS-CoV2 genome (42). Spike can be divided into S1 and S2 subunits based on the cleavage sites and functions (entry and fusion). S1 subunit involves in cellular entry, whereas the S2 subunit is required for membrane fusion. A Receptor Binding Domain (RBD)(amino acids (aa) 319-541) in the S1 subunit is responsible for binding to the ACE2 receptor (43). This attachment is crucial for viral entry leading to further steps in SARS-CoV2 replication. Mutations in the RBD domain are critical. Depending on mutations, the binding affinity of spike to ACE2 can change. Any mutations that favor higher attachment subsequently entry will be selected for further circulation and spread. Thus any mutations in SARS-CoV2 spike RBD, which allow higher binding to ACE2 receptor, will increase its ability to spread (44-50). The immune system generates antibody responses against spike proteins and antibodies against SARS-CoV-2 spike neutralize the virus and prevent infection. Interestingly, some studies have shown that convalescent sera are not as potent against some mutant version of the spike. Meanwhile, all emerging variants have different spike mutations, and many are linked to the higher transmission and reduce antibody neutralization (45-48, 51-57). Overall, because of its requirement for viral entry, the spike is a rapidly evolving protein at the center of the host-viral interface.

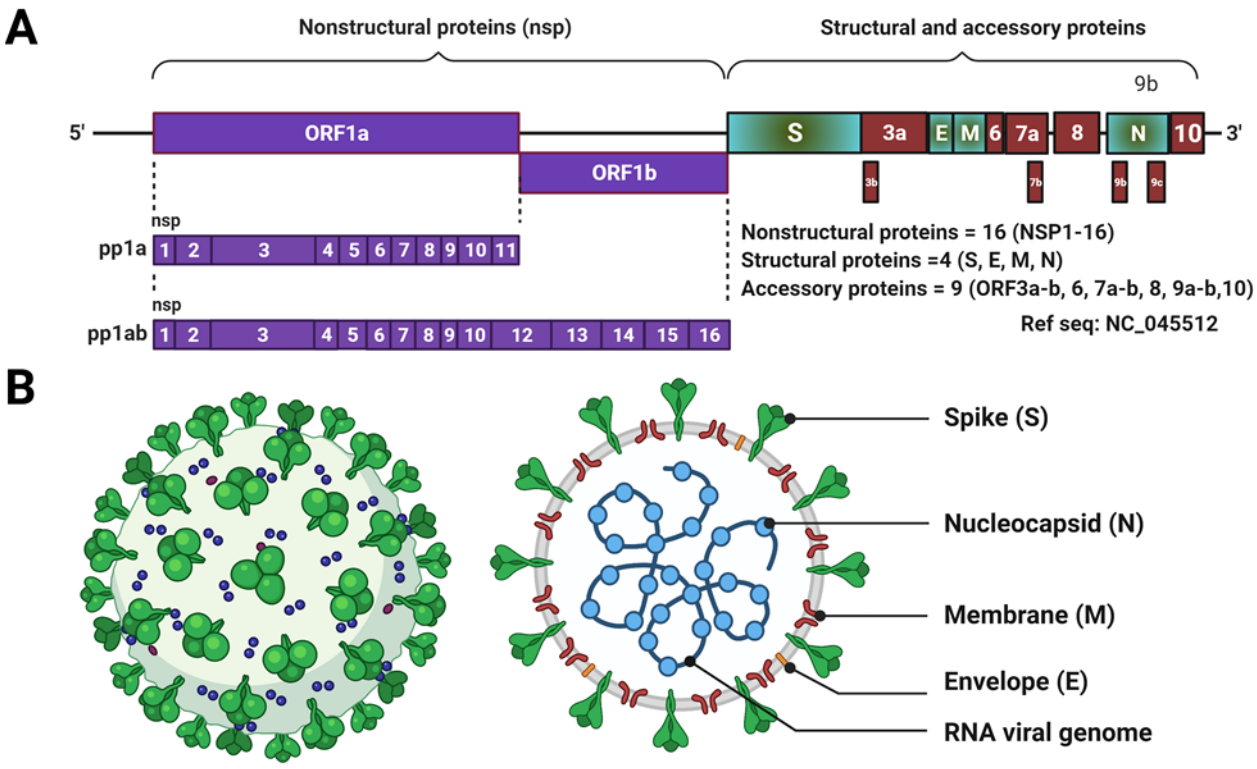


Figure 3: SARS-CoV2 genome and virion structure. **A.** SARS-CoV2 has two long polyproteins (pp1a and pp1ab) that encode 16 non-structural proteins from 5' end. From the 3' end of the genome encodes four structural proteins, namely spike (S), envelope (E), membrane protein (M), and nucleocapsid (N). Also, nine accessory proteins are predicted to be encoded. **B.** Virion structure showing spike forming a corona-like structure where RNA genome is protected inside surrounded by other structural proteins.

SARS-CoV-2 transmission and spread must be maintained for new spike mutations so that virus variants can be selected for a fitness advantage. The more infection that occurs, the higher the chance a new variant with a fitness advantage emerges. With the high number of infections within the 1.5 years of the pandemic, several variants have been observed in different parts of the world, namely from the United Kingdom(UK), United States of America (USA) (California (CA), and New York(NY)), India, Brazil, and South Africa (**Table 1** and **Figure 4**) (52, 53, 55-63). All these variants have the potential to escape from antibody-mediated immunity (escape mutants) and thus cause for concern. Thankfully, a recent report demonstrated that Pfizer vaccination elicits a protective response to virtually all known SARS-CoV2 variants. Vaccination gives better protection and higher antibody titer than the convalescent group and individuals infected naturally (64-68). Current vaccines selectively target spike protein but targeting multiple viral proteins might provide better protection. Targeting multiple viral proteins might reduce the chances of escape

mutants. Further investigation into the molecular biology of these variants will significantly benefit our understanding of viral evolution and help better vaccine design.

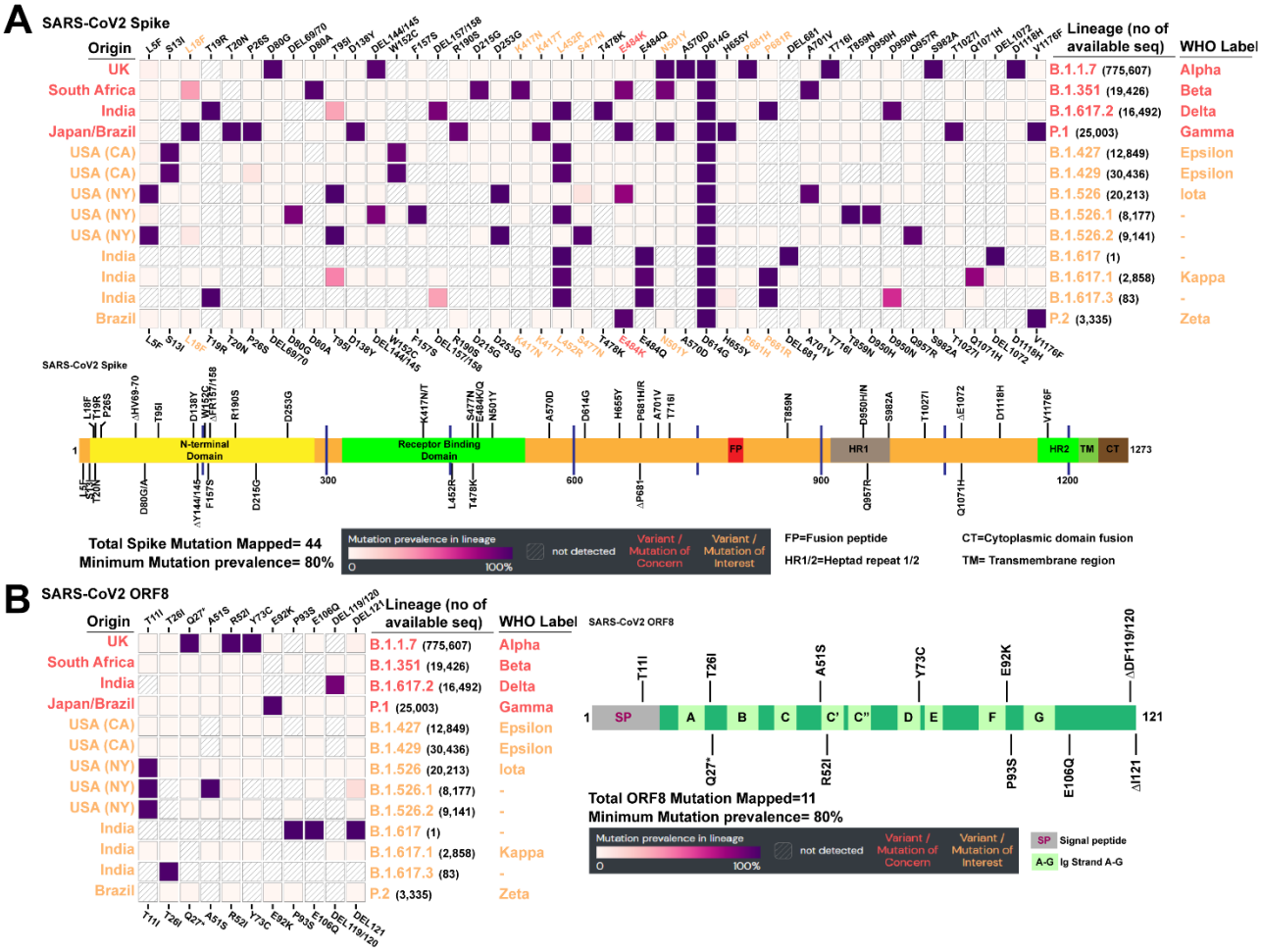


Figure 4: Mutation map of SARS-CoV2 spike and ORF8 among major variants. A. SARS-CoV2 Spike and **B.** SARS-CoV2 ORF8 are two hypervariable proteins prone to numerous mutations/deletions (Updated to 1st June 2021). Origin describes places where the variants were first reported, and lineages showed assigned names by the Centers for Disease Control and Prevention (CDC). In addition, the US Department of Health and Human Services (HHS) assigned naming variants of interest (VOI) and variants of concern (VOC) to distinguish between different variants transmission and severity. No. of the sequences shows the total available sequences were for comparison. The figure was adapted and edited from the outbreakinfo.com website (69).

1.2.1 SARS-CoV2 spike L452R mutation increased viral transmission, infectivity, and immune escape

SARS-CoV2 spike acquired many mutations in the ongoing COVID-19 pandemic. The emergence of multiples mutations, primarily in spike and sometimes other proteins, gave rise to variants with changes in receptor binding ability and immune evasion named Variant of Interest (VOI). Some variants that showed higher infectivity, higher transmission, and immune system evasion are called Variants of Concern (VOC). This nomenclature is established by the US Department of Health and Human Services (HHS). Later, to monitor and track different variants, WHO came with a nomenclature system to simplify with non-stigmatizing labels (**Table 1**) (70, 71). Among spike mutations, the most notable mutation is D614G. D614G was reported in early February 2020 and soon quickly became the most circulating variant in Europe (72). D614G is shown to enhance infectivity, which is believed to be explained by a higher open conformation state of spike protein in the D164G variant, which favors increased ACE2 binding resulting in the successful spread of this variant (44, 72-77). Later in May 2020, N501Y and E484K spike mutations appeared in the Beta variant, which also had D614G (**Table1** and **Figure 4**). N501Y has a higher affinity for ACE2 binding and increased transmissibility and soon appeared in other new variants (Alpha and Gamma) (56, 78, 79). Both UK and Brazil faced devastating effects from SARS-CoV2 infection from these variants. In the USA, new variants were reported with spike mutations in California (L452R) and New York (E484K) (55, 57, 58, 80). In this thesis, we report the transmission, infectivity, and neutralization of the Epsilon variant, which emerged in California, the USA (51, 55, 57, 80). Specifically, we reported increased transmissibility and infectivity of spike L452R mutation. We also observed that L452R mutation had reduced antibody neutralization in both convalescent patients and vaccine recipients. We found that N501Y mutation has higher infectivity than L452R in SARS-CoV2 pseudotyped virus entry assay in HEK293 cells that overexpression ACE2 and TMPRSS2 (HEK293-ACE2-TMPRSS2) cells and Human Airway Organoidsthat overexpress-ACE2 (HAO-ACE2) (57). Thus L452R mutation increased the transmissibility of the Epsilon variant.

L452R mutation has also been found in the Delta (VOC) and Kappa (VOI) variants (61). A newly reported A.27 variant in Germany does not have D614G mutation but has L452R spike mutation (60). Some of the same mutations present in G614-variants have appeared in those D614-variants suggesting convergent evolution. Studying the role of different spike mutations, such as L452R, on infectivity, transmission, and neutralization will provide insights into viral infectivity, transmission, and evolution. This knowledge is useful for vaccine design and future preventive strategies (51, 81).

Table 1: **SARS-CoV2 variants: VOC (Variants of Concern) and VOI (Variants of interest)**.(70, 71)

	Lineage/Clade/label				First appearance	
	WHO	**GISAID**	**Next strain**	Pango	**Country**	**Year**
Variants of Concern	Alpha	GRY*	20I/S:501Y.V1	B.1.1.7	UK	09/2020
	Beta	GH/501Y.V2	20H/S:501Y.V2	B.1.351	South Africa	05/2020
	Gamma	GR/501Y.V3	20J/S:501Y.V3	P.1	Brazil	11/2020
	Delta	G/452R.V3	21A/S:478K	B.1.617.2	India	10/2020
Variants of Interest	Epsilon	GH/452R.V1	20C/S.452R	B.1.427/ B.1.429	USA (CA)	04/2020
	Zeta	GR	20B/S.484K	P.2	Brazil	03/2020
	Eta	G/484K.V3	20A/S484K	B.1.525	Multiple	12/2020
	Theta	GR	20B/S:265C	P.3	Philippines	01/2021
	Iota	GH	20C/S:484K	B.1.526	USA (NY)	11/2020
	Kappa	G/452R.V3	21A/S:154	B.1.617.1	India	10/2020

*(formerly GR/501Y.V1)

1.2.2 Multi-functional ORF8 is an accessory protein with immunomodulatory property

Besides structural proteins and non-structural proteins, SARS-CoV2 also encodes accessory proteins (**Figure 3**). Although several accessory proteins are suspected to be encoded based on open reading frames, only a few were found to encode proteins (82). SARS-CoV2 ORF8 is a unique accessory protein in the betacoronavirus group. ORF8 shows strong evolutionary evidence for protein-encoding (82). ORF8 is a 121aa long protein, prone to mutations and deletions (83). Along with spike, it is one of the most hypervariable regions in the SARS-CoV-2 genome, suggesting it might be involved in viral-host interactions (84). Deletion (Δ382nt in SARS-CoV2 genome caused almost complete ORF8 deletion) and stop mutations that prevent ORF8 production are found in replication-competent SARS-CoV2, suggesting that ORF8 might not be involved in viral replication and could be dispensable (73). In COVID-19 patients, anti-ORF8 antibodies are readily detectable and have been suggested as a serological detection marker for COVID-19 (85, 86). A report indicates that ORF8 is primarily detected by IgM antibodies (85). Thus, the ORF8 protein is an interesting protein with a possible role in viral evolution, pathogenesis, and diagnostic purposes (87).

Structural analysis shows evidence for SARS-CoV2 ORF8 as a viral protein with a unique immune evasion function compared to SARS-COV ORF8 (88). Sequence homology analysis

shows SARS-CoV2 ORF8 is more similar to bat and pangolin ORF8 than that of SARS-CoV and MERS-CoV (83). Although SARS-CoV ORF8 protein also experienced similar deletions and mutations and later emerged as ORF8a and ORF8b, SARS-CoV2 ORF8 is still a single ORF8 protein (89). SARS-ORF8a and ORF8b are suspected of having resulted from the founder effect, although it is unknown if this is true for SARS-COV2 (90). ORF8 can form dimers and larger assembly. A specific domain, (73YIDI76) unique to SARS-CoV2, is predicted to be responsible for dimer stabilization and larger assembly (88). ORF8 has an N-terminal signal peptide (aa1-15) domain followed by multiple immunoglobulins (Ig)-like folds (**Figure 4B**). The presence of signal peptides suggests that ORF8 probably gets secreted through the ER-Golgi secretion pathway (83, 88, 91). Secreted SARS-CoV2 ORF8 has been detected in cell culture media and infected patient sera (86). The presence of ORF8 antibodies, secreted form of ORF8 protein, and variations in ORF8 sequence make ORF8 an interesting protein for further research.

Multiple mutations and deletions have been found in the ORF8 protein (**Figure 4B**). ORF8 is probably involved in the fitness of viruses in non-human hosts (82). ORF8 is the least conserved area in the SARS-CoV2 genome. ORF8 is almost entirely deleted in many clinical isolates identified in Bangladesh, Taiwan, Australia, and Singapore, where these isolates have been linked to reduced severity (63, 92, 93). A closer look at mutational frequency at ORF8 showed a possibility of a link to viral fitness. The L84S and S24L mutations have been linked to possible protein stability and immunogenicity (91). The mutation Y73C specific to the 73YIDI76 domain has been reported in the Alpha variant, but its effect is still unknown (**Table 1** and **Figure 4B**). Combinations of ORF8 mutations are observed in many clinical isolates (94). ORF8 mutations are often accompanied by other spike mutations (**Figure 4A** and **B**). Mutations in spike and other viral proteins, including ORF8, in combination probably play a significant role in viral transmission and pathogenesis. One study suggested that spike mutations resulting in increased cellular entry and transmission might be accompanied by ORF8 mutations that lessen individual disease severity. This might explain why the ORF8 region prone to deletion or mutations leading to non-functional protein (82). One way to think about this is that ORF8 might be more important for individual spread than population-level transmission. SARS-CoV2 ORF8 deleted variant showed reduced disease severity and hospital admittance (93). If patients have increased survival and less hospital admittance, they have higher chances of spreading the virus. Another explanation relates to the adaptability of the virus between two hosts (bat or pangolin and human). The presence of almost identical (90-96%) ORF8 in bat and pangolin suggests that the virus probably needed this protein for circulation in those hosts, which have very different immune responses from humans (83). Constant mutations suggest that ORF8 is under positive selection pressure to

become less immunogenic in the host to help with viral replications (73, 84). Thus ORF8 is a protein with potential roles in SARS-CoV2 transmission and host adaptation.

ORF8 interacts with numerous host proteins suggesting it can disrupt or dysregulate host proteins or pathways. ORF8 interacts or dysregulates host proteins or pathways linked to host defense and immunity. Proteomics data show it interacts with matrix metalloproteases which are linked to immunopathogenesis (95, 96). It downregulates major histocompatibility complex class I (MHC-I) expression and interferon (IFN) activation (97, 98). At the same time, ORF8 is shown to dysregulate transforming growth factor-beta (TGF-β) protein function (95). Structural analysis shows ORF8 might be involved in the immune evasion pathway to facilitate viral replication (88). This thesis showed that ORF8 interacts with the Interleukin 17 Receptor A (IL17RA) pathway. One study showed ORF8 interaction with IL17RA has a role in IL17 mediated cytokine storm signaling (99). These pathways have been linked to inflammation and alveolar damages observed in COVID-19 patients. These interactions dysregulate normal immunomodulatory functions to increase viral replication and spread. Targeting ORF8 is probably problematic due to hypervariability in that region, but understanding which pathway ORF8 affects might give us opportunities to use inhibitors or activators that might limit disease severity and viral spread. The function and mechanisms of ORF8 are still a mystery and whether SARS-CoV2 needs ORF8 is still under debate.

1.2.3 IL17RA involves in SARS-CoV2 replication and pathogenesis

IL17RA is a member of the Interleukin 17 Receptor (IL17R) family, including four other members (Interleukin 17 Receptor B (IL17RB), Interleukin 17 Receptor C (IL17RC), Interleukin 17 Receptor D (IL17RD) and Interleukin 17 Receptor E (IL17RE)). Interleukin 17 A (IL17 A) and Interleukin 17 F (IL17F) cytokines form homo or heterodimers and bind to IL17RA and IL17RC subunits (100). IL17RA can also form a complex with other receptors like IL17RB, IL17RE and transduce IL17C and IL17E mediated signaling (101). IL17RA is expressed by most of the cells, while very few cell types express IL17RC. IL17RA has an extracellular domain, a transmembrane domain (TM), and an intracellular domain (101, 102). An isoform that lacks the TM domain named sIL17RA (soluble IL17RA) gets secreted and involved in IL17RA signaling (103, 104). IL17A cytokine driven signaling, which requires IL17RA receptors, is secreted by T helper 17 cells (Th17). After binding cytokines, IL17RA mediated signaling induces nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), mitogen-activated protein kinase (MAPKs), and CCAAT-enhancer-binding proteins (C/EBP) pathways that activate target gene activation, leading to inflammation (101). Thus, IL17A and IL17RA are very important for inflammation, pulmonary

defense, chronic respiratory illness, and auto-immunity (105, 106). The role of IL17A and IL17RA can be yin and yang, depending on the expression and mechanism they play (107). Moreover, targeting IL17A and IL17RA using monoclonal antibodies alleviated immune-mediated disease (108, 109).

IL17RA signaling has been shown to be important in lung injury and infection (105, 107, 110-113). However, relatively little is known about the role of IL17 signaling during viral infection. In one study, knockout of IL17RA in mice showed to reduce neutrophils recruitment to the lung during Influenza infection resulting in reduced lung injury. The same research showed that Influenza infection caused increased expression of IL17A (102). A similar phenomenon was observed in Herpes simplex virus-1 (HSV-1) infection, where IL17RA KO mice have reduced corneal inflammation (114). In the Iranian population, severe pneumonia has been associated with IL17A SNPs (115). IL17A has been shown to inhibit Respiratory Syncytial Virus (RSV) clearance by increasing mucus production (116). In Simian Immunodeficiency Virus (SIV) infections, it down-regulates IL17A expression by decreasing the Th17 cells population. In IL17RA KO mice, increased bacterial infection was observed (117). Influenza A is shown to inhibit Interleukin 1 beta (IL-1β) production in the co-infection model targeting IL17RA signaling (118). In DENV infection, IL17RA knockout mice showed reduced fatality during viral infection (119). The role of IL17A has also been documented in Chikungunya Virus (CHIKV) infection (120). These show that viral infections can result in dysregulations of the IL17RA pathway leading to viral pathogenesis.

In severe COVID-19 patients, elevated cytokine levels have been observed that can cause cytokine storm, which is the overproduction of cytokines that can lead to organ damage. In these patients, increased or dysfunctional IL17 expression or signaling was reported (121, 122). The IL17 target genes such as IL6 were found to be upregulated in severe patients. IL6 can contribute to lung inflammation and cytokine storm. Therefore dysfunctional IL17RA signaling can increase the severity in COVID-19 (121, 122). A genome-wide association study (GWAS) in COVID-19 patients found IL17A alleles associated with disease severity (123). In mild COVID-19 patients, sIL17RA was increased, suggesting internalization and decreased IL17RA signaling by sIL17RA is beneficial (124, 125). This thesis showed that SARS-CoV2 ORF8 interacts with endogenous IL17RA, and IL17RA knockdown decreased SARS-CoV2 replication (124). Given that ORF8 protein can be secreted and dysregulate IL17RA signaling, IL17RA is an interesting pathway to target and study for SARS-CoV2 infection (87, 99). IL17RA signaling is also crucial in suppressing bacterial and fungal infections (111, 126). Recently, bacterial, fungal, and viral co-infections are

found in COVID-19 patients suggesting that immunity against these pathogens is weakened or dysfunctional (127-129). This may be linked to IL17RA signaling disruptions or dysfunction by SARS-CoV2. Besides acute infection, little is known about why many COVID-19 patients carry 'long-COVID' symptoms where symptoms persist long even the acute phase of the disease has passed (130-133). It is still unknown whether dysregulations of the IL17RA pathway have any role in co-infections and long-COVID symptoms. Targeting IL17RA or IL17A to reduce COVID-19 severity can be beneficial (134, 135). SARS-CoV2 uniquely targets the IL17RA pathway, suggesting that this pathway has a prominent role in SARS-CoV2 transmission and pathogenesis.

1.2.4 The gap in knowledge about SARS-CoV2 Spike and ORF8 protein

Continuous viral mutations in large populations worldwide give rise to variants with different transmission and antibody neutralization characteristics. We need to identify which variants pose threats quickly by analyzing their characteristics. These characteristics can be answered by using various tools and techniques. The SARS-CoV2 spike pseudotyped virus assay technique is a lentiviral system with SARS-CoV2 spike rather than traditional envelope proteins like VSV-G. It is a popular and easy tool for viral entry assay, although it usually needs over-expression of SARS-CoV2 entry receptors hACE2. SARS-CoV2 pseudotyped virus does not have other SARS-CoV2 structural proteins like Virus-like Particles (VLP), Replicons (infectious or non-infectious), and the full infectious virus. Thus, often a complete picture or combined effects of structural protein mutations are sometimes missed. At the same time, the SARS-Cov2 spike pseudotyped virus provides a clean and entry-only assay that can be beneficial to studies different mutation entry and neutralization. If these studies are done along with infectious clones or whole viruses, a complete mechanism of viral entry and transmission can be understood. However, with new variants emerging, the infectivity of lots of the spike mutations is unknown. New assay and tools need to be discovered to give quick and reproducible data. At the same time, combining real world data with mechanisms can lead us to preventive measures.

The SARS-CoV2 infection leads to non-symptomatic to severe fetal conditions and sometimes long COVID. In severe COVID-19 patients, IL17RA pathway-induced gene expression is found to be upregulated. ORF8 interacts with many other host factors where the consequence of these interactions is unknown. Very little is known about ORF8 biology regarding its pro-viral role in viral transmission and fitness. The secretory version of ORF8 needs to be investigated. Whether ORF8 is responsible for long COVID symptoms is unknown. Investigation on the role of ORF8 biology in viral fitness and host-adaptation would help to understand SARS-CoV2 pathogenesis.

1.3 ZIKV endemic raises concerns about more vector-mediated outbreaks

ZIKV virus is a flavivirus of the *Flaviviridae* family. *Flaviviridae* family includes commonly known viruses like Dengue Virus (DENV), CHIKV, Yellow Fever Virus (YFV), and West Nile Virus (WNV), and many of them spread by mosquitoes. *Flavivirus* genus contains 53 identified viruses. Some viruses in that genus, like DENV and YFV, can cause severe illnesses like Viral hemorrhagic fever, while some, like WNV, can cause severe illness, encephalitis, and death (136). Like these viruses, ZIKV has evolved to adapt both host and vector. The commensal microbiome in mosquito saliva and initial interaction at host skin plays an essential role in viral replication (137). ZIKV can transmit by the mosquito, sexual contact, mother to child, and rarely by blood transfusion. ZIKV proteins are evolved to interact and target both human and mosquito proteins and pathways (138). With viral evolution, ZIKV acquires mutations over time that change their characteristics. ZIKV has two major lineages (Asian and African), which show different infectivity and pathogenesis (139). Although ZIKV was identified a long time ago, little research has been done on the molecular biology of ZIKV.

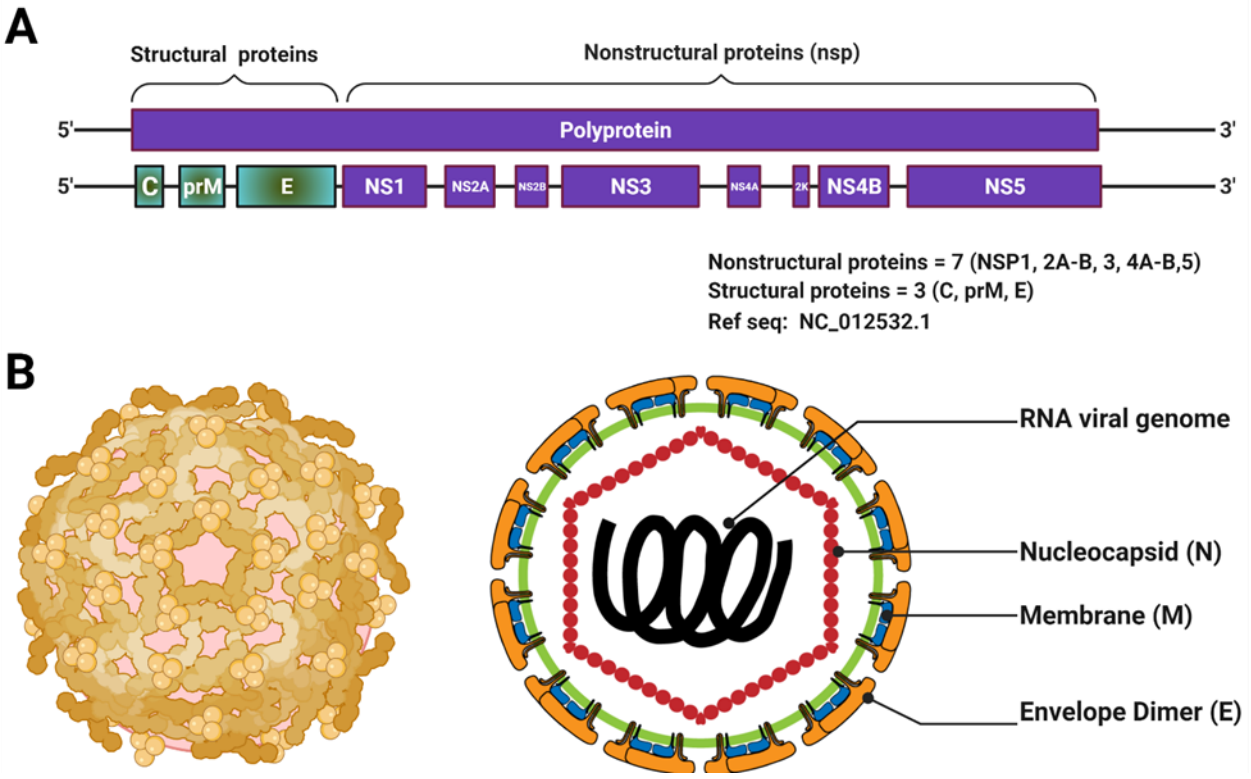


Figure 5: ZIKV genome and virion structure. A. ZIKV encodes a long polyprotein that is proteolytically processed and later encodes seven non-structural proteins and three structural proteins. Three structural proteins are capsid (C), envelope (E), and membrane protein (prM). **B.**

Virion structure showing envelope dimer with membrane protein surrounding RNA genome protected by capsid.

ZIKV has a +ssRNA genome with a length of ~11kb. ZIKV genome encodes a long polyprotein that is proteolytically processed to seven non-structural and three structural proteins (**Figure 5**). ZIKV structural proteins are primarily involved in viral entry, genome release, packaging, and release of viral particles. In contrast, non-structural proteins attributed to RNA replications, immune evasion, and host cellular pathway disruptions (140, 141). ZIKV genome starts its translations at 5' end, which encodes three structural proteins: capsid (C), membrane (prM), and envelope (E) (**Figure 5**). The capsid is the first protein encoded by ZIKV polyprotein. ZIKV capsid is 122aa long, later processed, and cleaved to functional 104aa long protein (142, 143). Uncleaved ZIKV capsid is termed as 'anchored capsid'. The capsid is a structural protein with two main functions: the release of viral RNA during uncoating and recruiting viral genomic RNA during assembly (144, 145). Mature infectious viral particles cannot be produced without capsid even though prM and E can form a subviral particle (146, 147). Capsid also packages viral RNA during virion formation (**Figure 5 B**) (148). ZIKV capsid interacts with multiple host proteins involving translation, RNA processing, and developments(138). Thus ZIKV capsid has significant roles in viral life cycle and pathogenesis.

1.3.1 Multifaceted functions of ZIKV capsid attributed to advantages for viral replication

Structural analysis shows the positively charged motif of ZIKV capsid responsible for capsid binding with viral genome. On the other hand, the side of the hydrophobic cleft of ZIKV capsid interacts with lipid bilayers, endoplasmic reticulum, or viral envelope. These interactions help during viral assembly and viral RNA release during uncoating (149). Even though flavivirus can form empty viral particles without capsid, the capsid is required for an infectious particle formation. ZIKV structure shows prM and E protein form the outer icosahedral shell (150, 151). In immature ZIKV particles, it was demonstrated that residual ZIKV capsid inner shells in contact with the inner viral membrane. Mature virus lacking this residual shell contact suggests capsid went through rearrangement during maturation (150). ZIKV Capsid has been shown to have broad binding to nucleotide types (ssDNA, dsDNA, ssRNA, dsRNA). So, capsid protein might involve in DNA mediated host response (152). However, the role of ZIKV capsid binding to different nucleotide types is still unknown.

Very little is known about the ZIKV capsid and its function. Besides its role in viral replication, it has been shown to interacts with multiple nuclear and nucleolar host proteins and

targeting RNA surveillance pathways (138, 153-156). ZIKV Capsid is showed to interact with lipid droplets which helps the ZIKV life cycle (152, 157). ZIKV capsid also inhibits stress granule formation and hijack stress granule factors for viral replication (158). ZIKV capsid with several other ZIKV proteins has been shown to inhibits cell growth and cause cell-cycle dysregulations. ZIKV capsid causes cell cycle arrest and cell deaths (159). In another study, ZIKV capsid targets the microRNA (miRNA) biogenesis pathway by interacting with Dicer, a key enzyme of miRNA biogenesis. ZIKV capsid also inhibits Dicer activity disrupting normal neurogenesis, thus can lead to neurodevelopment defects (160). Nucleolar localization of ZIKV capsid has been shown to cause ribosomal stress and apoptosis in neurons, suggesting neurodevelopmental disruption (161). This thesis showed that ZIKA capsid targets host RNA surveillance pathway (NMD) and downregulate nuclear UPF1 (155). We also showed that ZIKV infection caused the accumulation of host mRNA in the nucleus in human neural progenitor cells (NPC), as observed in UPF1 knockdown. This accumulation results in trapping FRAS1 Related Extracellular Matrix 2 (*FREM2)* transcript, an essential protein for NPC differentiation, resulting in disruption of neurodevelopment but enhancing ZIKV replication (162). This suggests ZIKV capsid has a unique role in the nucleus, leading to a neurodevelopmental defect.

1.3.2 NMD is an RNA surveillance pathway that restricts ZIKV replication

The non-sense mediated RNA decay (NMD) pathway is a cellular RNA surveillance pathway that targets RNA transcripts with premature termination codons(PTC) and aberrant mRNA. NMD pathway can either exon-junction complex (EJC) dependent or independent mediated RNA degradation. In the EJC dependent pathway, EJC loaded on spliced mRNA with PTC cannot be displaced. This leads to activation of NMD pathway recruiting Up-frameshift suppressor 1 homolog UPF1) is the master regulator of NMD pathway. Phosphorylation of UPF1 after recruitment leads to recycling of EJC, and transcripts became subjected to endo or exo-nucleolytic decay leading to degradation of aberrant transcript (163). The NMD pathway restricts RNA viral transcription by targeting its RNA transcripts, thus acts as a restriction factor for RNA viruses and retroviruses. RNA viruses employ viral proteins that bind to different NMD pathway members and disrupt their function to avoid and evade host response (163). Also, viral infection in the cell downregulates the NMD pathway. Downregulation and knockdown of NMD members lead to an increase in viral replication. Flaviviruses are known to target NMD pathways (164). The flaviviral capsid has been reported to target NMD members. Flaviviral interactome networks show flaviviral proteins interact with multiple members of the NMD pathway (138, 164). ZIKV capsid targets nuclear UPF1 and mimics neurodevelopmental defects (155, 165). Furthermore, UPF1

plays an essential role in neural and development differentiation (166, 167). Thus effect of ZIKV infection on UPF1 leading to neurological defects is an interesting area of investigation.

In this thesis, we showed, ZIKV Capsid expression and infection decrease nuclear UPF1 expression. Knockdown of UPF1 caused increased viral replication (155). We also showed that ZIKV infection also prevents the export of host mRNA from the nucleus to the cytoplasm (162). This result indicates that ZIKV specifically targets the NMD pathway, decreases UPF1 expression, and disrupts regular functions of UPF1, which helps ZIKV replication. Even ZIKV replicates in the cytoplasm, ZIKV localizes to the nucleus to suppress the host RNA surveillance pathway by targeting UPF1. The reason for trapping host mRNA in the nucleus is probably to reduce the amount of mRNA transcript available for translation in the cytoplasm. This allows host cellular translation machinery at the highest usage for viral RNA translation, thus enhancing viral production. This thesis showed that ZIKV infection caused the trapping of *FREM2* transcripts, resulting in premature neuronal differentiation (162). This premature differentiation might lead to fewer brain cells number, thus reducing the fetus's brain size, which is a characteristic feature in fetal microcephaly. It has been shown that ZIKV infection causes premature differentiation and NPC depletion in brain organoids (39, 168). Suppression of UPF1 expression and function enhances ZIKV replication. This suggests UPF1 is a restriction factor for ZIKV. Thus ZIKV capsid has a unique role in nuclear UPF1 downregulation for viral replication and pathogenesis.

1.3.3 The gap in knowledge in ZIKV Capsid functions

There is very little research on the role and functions of ZIKV capsid. ZIKV capsid has a prominent role in viral replication and assembly. Broad binding abilities to different nucleotides and interactions with many nuclear proteins give rise to possibilities of a more prominent role of ZIKV capsid, specifically in the nuclear environment. Disrupting nuclear proteins' function allows ZIKV to replicate better in the cytoplasm. ZIKV capsid evolved to interact with multiple proteins in the cytoplasm and in the nucleus to perturb pathways for enhancing viral replication and spread. Whether ZIKV capsid can binds to host DNA is still unknown. It would be interesting to see what other nuclear proteins interacts with ZIKV capsid and how it affects viral replication. At the same time, by understanding this interaction, we would be able to comprehend ZIKV pathogenesis.

1.4 Deciphering host-viral interactions using proteomics

Proteomics is a tool that allows knowing how different proteins interact with each other inside a cell. Understanding these interactions leads to a better understanding of the host pathway, response, and diseases (169, 170). Targeting and treating an infectious viral disease by the drug

can be done in two ways: targeting the virus and targeting host proteins or pathways required for viral replication, pathogenesis, and transmission. Host-pathogen interactome networks show the cellular proteins targeted by the virus. If we know which pathways are essential for viral replication or pathogenesis, we can target those and find a cure or improve preventive measures. Thus if we can decipher host-pathogen interactions, we can quickly narrow down the targets that can be used against the virus (40, 95, 124, 138, 153, 154, 169, 171, 172). In a pandemic scenario, we can use this knowledge to repurpose the previously approved drug to limit the spread. This has been done in the SARS-CoV2 pandemic (40, 95, 124, 171). This saves money and time when an urgent need for preventive measures needed in an outbreak. Understanding these interactions also allows finding targets that causing disease complications and symptoms.

1.5 Scope and outline of the thesis

This thesis showed research on viral proteins of the current SARS-CoV2 pandemic and recent ZIKV epidemic. The current ongoing SARS-CoV2 pandemic has taken millions of lives and caused a worldwide economic crisis. In the case of the ZIKV, the epidemic caused increased fetal microcephaly, which is a developmental defect. SARS-CoV2 has zoonotic transmission, whereas ZIKV has vector-borne transmission. Both of these viruses have similar viral replication and targets host cellular pathways and proteins. The recent and current history of both viruses shows how these viruses have the potential for more outbreaks, epidemics even pandemics. This is why more emphasis on understanding the molecular biology of these viruses are needed,

Chapter 2 describes the identification of a new SARS-CoV2 variant Epsilon (B.1.427/B.1.429). This chapter demonstrates the variant's role in infectivity, transmission, and antibody neutralization. **Chapter 3** focuses on identifying conserved and shared mechanisms for SARS-CoV2 using host-coronavirus protein-protein interaction network and functional genetics. Four targets were identified for SARS-CoV2 and validated using in-vitro experiments and real-world data. In **Chapter 4**, ZIKV capsid's role in downregulating NMD pathway and effect on viral replication was described. **Chapter 5** describes the effect of UPF1 downregulation by ZIKA infection on nuclear mRNA transport. It also identified a gene in which mRNA export is perturbed and its function, leading to altered differentiation in NPC. Finally, **chapter 6** summarizes the essential findings and future perspectives.

References

1. McCollum, A.M., et al., *Poxvirus viability and signatures in historical relics.* Emerg Infect Dis, 2014. **20**(2): p. 177-84.
2. Fenner, F., et al., *Smallpox and its eradication / F. Fenner ... [et al.].* 1988, World Health Organization: Geneva.
3. Hopkins, D.R., *Smallpox: ten years gone.* American journal of public health, 1988. **78**(12): p. 1589-1595.
4. Flight, C., *Smallpox: Eradicating the Scourge*, in *BBC*. 2011, BBC.
5. Power, A.G. and C.E. Mitchell, *Pathogen spillover in disease epidemics.* Am Nat, 2004. **164 Suppl 5**: p. S79-89.
6. Daszak, P., A.A. Cunningham, and A.D. Hyatt, *Emerging infectious diseases of wildlife--threats to biodiversity and human health.* Science, 2000. **287**(5452): p. 443-9.
7. WHO. *WHO information on tuberculosis and pandemic influenza A (H1N1)*. 2009 [cited 2021; Available from: https://www.who.int/news/item/12-11-2009-who-information-on-tuberculosis-and-pandemic-influenza-a-(h1n1).
8. WHO. *Coronavirus*. [cited 2021; Available from: https://www.who.int/health-topics/coronavirus.
9. CDC. *Zika Virus*. 2019 [cited 2021; Available from: https://www.cdc.gov/zika.
10. Li, W., et al., *Bats are natural reservoirs of SARS-like coronaviruses.* Science, 2005. **310**(5748): p. 676-9.
11. Wu, F., et al., *A new coronavirus associated with human respiratory disease in China.* Nature, 2020. **579**(7798): p. 265-269.
12. Zhang, T., Q. Wu, and Z. Zhang, *Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak.* Curr Biol, 2020. **30**(7): p. 1346-1351 e2.
13. WHO. *The history of zika virus*. 2016 [cited 2021; Available from: https://www.who.int/news-room/feature-stories/detail/the-history-of-zika-virus.
14. Zanluca, C., et al., *First report of autochthonous transmission of Zika virus in Brazil.* Mem Inst Oswaldo Cruz, 2015. **110**(4): p. 569-72.
15. Campos, G.S., A.C. Bandeira, and S.I. Sardi, *Zika Virus Outbreak, Bahia, Brazil.* Emerg Infect Dis, 2015. **21**(10): p. 1885-6.
16. CDC. *COVID-19*. Available from: https://www.cdc.gov/coronavirus/2019-nCoV.

17. Ji, W., *Coronaviruses: What Should We Know About the Characteristics of Viruses?*, in *Coronavirus Disease - COVID-19*, N. Rezaei, Editor. 2021, Springer International Publishing: Cham. p. 23-39.
18. Pierson, T.C. and M.S. Diamond, *The continued threat of emerging flaviviruses.* Nat Microbiol, 2020. **5**(6): p. 796-812.
19. McBroom, K., *A Comparison of Zika Virus and COVID-19: Clinical Overview and Public Health Messaging.* J Midwifery Womens Health, 2021.
20. Dong, E., H. Du, and L. Gardner, *An interactive web-based dashboard to track COVID-19 in real time.* Lancet Infect Dis, 2020. **20**(5): p. 533-534.
21. Mavrodiev, E.V., et al., *On Classification and Taxonomy of Coronaviruses (Riboviria, Nidovirales, Coronaviridae) with special focus on severe acute respiratory syndrome-related coronavirus 2 (SARS-Cov-2).* bioRxiv, 2020: p. 2020.10.17.343749.
22. Cui, J., F. Li, and Z.L. Shi, *Origin and evolution of pathogenic coronaviruses.* Nat Rev Microbiol, 2019. **17**(3): p. 181-192.
23. Singh, D. and S.V. Yi, *On the origin and evolution of SARS-CoV-2.* Exp Mol Med, 2021. **53**(4): p. 537-547.
24. Zhou, H., et al., *Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses.* bioRxiv, 2021: p. 2021.03.08.434390.
25. Mallapaty, A.M.S. *The COVID lab-leak hypothesis: what scientists do and don't know*. 2021.
26. Zhou, P., et al., *A pneumonia outbreak associated with a new coronavirus of probable bat origin.* Nature, 2020. **579**(7798): p. 270-273.
27. Greenhalgh, T., et al., *Ten scientific reasons in support of airborne transmission of SARS-CoV-2.* Lancet, 2021. **397**(10285): p. 1603-1605.
28. Gibson, P.G., L. Qin, and S.H. Puah, *COVID-19 acute respiratory distress syndrome (ARDS): clinical features and differences from typical pre-COVID-19 ARDS.* Med J Aust, 2020. **213**(2): p. 54-56 e1.
29. Valentine, M.J., C.C. Murdock, and P.J. Kelly, *Sylvatic cycles of arboviruses in non-human primates.* Parasit Vectors, 2019. **12**(1): p. 463.
30. Agboli, E., et al., *Mosquito-Associated Viruses and Their Related Mosquitoes in West Africa.* Viruses, 2021. **13**(5).
31. Diallo, M., et al., *Amplification of the sylvatic cycle of dengue virus type 2, Senegal, 1999-2000: entomologic findings and epidemiologic considerations.* Emerg Infect Dis, 2003. **9**(3): p. 362-7.

32. Masmejan, S., et al., *Zika Virus.* Pathogens, 2020. **9**(11).
33. Rubin, E.J., M.F. Greene, and L.R. Baden, *Zika Virus and Microcephaly.* N Engl J Med, 2016. **374**(10): p. 984-5.
34. Stower, H., *Zika virus shedding in semen.* Nat Med, 2018. **24**(6): p. 702.
35. Dick, G.W., S.F. Kitchen, and A.J. Haddow, *Zika virus. I. Isolations and serological specificity.* Trans R Soc Trop Med Hyg, 1952. **46**(5): p. 509-20.
36. Koike, H., A. Chiba, and M. Katsuno, *Emerging Infection, Vaccination, and Guillain–Barré Syndrome: A Review.* Neurology and Therapy, 2021.
37. Teixeira, M.G., et al., *The Epidemic of Zika Virus-Related Microcephaly in Brazil: Detection, Control, Etiology, and Future Scenarios.* Am J Public Health, 2016. **106**(4): p. 601-5.
38. Khaiboullina, S.F., et al., *Zika Virus Transmission Through Blood Tissue Barriers.* Front Microbiol, 2019. **10**: p. 1465.
39. Gabriel, E., et al., *Recent Zika Virus Isolates Induce Premature Differentiation of Neural Progenitors in Human Brain Organoids.* Cell Stem Cell, 2017. **20**(3): p. 397-406 e5.
40. Gordon, D.E., et al., *A SARS-CoV-2 protein interaction map reveals targets for drug repurposing.* Nature, 2020. **583**(7816): p. 459-468.
41. Hoffmann, M., et al., *SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor.* Cell, 2020. **181**(2): p. 271-280 e8.
42. Perrotta, F., et al., *Severe respiratory SARS-CoV2 infection: Does ACE2 receptor matter?* Respir Med, 2020. **168**: p. 105996.
43. Wrapp, D., et al., *Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation.* Science, 2020. **367**(6483): p. 1260-1263.
44. Plante, J.A., et al., *Spike mutation D614G alters SARS-CoV-2 fitness.* Nature, 2021. **592**(7852): p. 116-121.
45. Cele, S., et al., *Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma.* Nature, 2021. **593**(7857): p. 142-146.
46. Wibmer, C.K., et al., *SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma.* Nat Med, 2021. **27**(4): p. 622-625.
47. Liu, Z., et al., *Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization.* Cell Host Microbe, 2021. **29**(3): p. 477-488 e4.
48. Xie, X., et al., *Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera.* Nat Med, 2021. **27**(4): p. 620-621.

49. Colson, P., et al., *Spreading of a new SARS-CoV-2 N501Y spike variant in a new lineage.* Clin Microbiol Infect, 2021.
50. Winger, A. and T. Caspari, *The Spike of Concern-The Novel Variants of SARS-CoV-2.* Viruses, 2021. **13**(6).
51. Focosi, D. and F. Maggi, *Neutralising antibody escape of SARS-CoV-2 spike protein: Risk assessment for antibody-based Covid-19 therapeutics and vaccines.* Rev Med Virol, 2021.
52. Liu, Y., et al., *The N501Y spike substitution enhances SARS-CoV-2 transmission.* bioRxiv, 2021.
53. Olsen, R.J., et al., *Identification and trajectory of growth of concerning SARS-CoV-2 variants in Houston, Texas, January through April 2021 based on 11,568 genome sequences.* medRxiv, 2021: p. 2021.05.20.21257552.
54. Bugembe, D.L., et al., *A SARS-CoV-2 lineage A variant (A.23.1) with altered spike has emerged and is dominating the current Uganda epidemic.* medRxiv, 2021: p. 2021.02.08.21251393.
55. Zhang, W., et al., *Emergence of a Novel SARS-CoV-2 Variant in Southern California.* JAMA, 2021. **325**(13): p. 1324-1326.
56. Tegally, H., et al., *Detection of a SARS-CoV-2 variant of concern in South Africa.* Nature, 2021. **592**(7854): p. 438-443.
57. Deng, X., et al., *Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant.* Cell, 2021.
58. Annavajhala, M.K., et al., *A Novel SARS-CoV-2 Variant of Concern, B.1.526, Identified in New York.* medRxiv, 2021.
59. Ozer, E.A., et al., *High prevalence of SARS-CoV-2 B.1.1.7 (UK variant) and the novel B.1.5.2.5 lineage in Oyo State, Nigeria.* medRxiv, 2021.
60. Mallm, J.-P., et al., *Local emergence and decline of a SARS-CoV-2 variant with mutations L452R and N501Y in the spike protein.* medRxiv, 2021: p. 2021.04.27.21254849.
61. Cherian, S., et al., *Convergent evolution of SARS-CoV-2 spike mutations, L452R, E484Q and P681R, in the second wave of COVID-19 in Maharashtra, India.* bioRxiv, 2021: p. 2021.04.22.440932.
62. van Dorp, L., et al., *Emergence of genomic diversity and recurrent mutations in SARS-CoV-2.* Infect Genet Evol, 2020. **83**: p. 104351.
63. Islam, M.R., et al., *Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity.* Sci Rep, 2020. **10**(1): p. 14004.

64. Sahin, U., et al., *BNT162b2 vaccine induces neutralizing antibodies and poly-specific T cells in humans.* Nature, 2021.
65. Frenck, R.W., Jr., et al., *Safety, Immunogenicity, and Efficacy of the BNT162b2 Covid-19 Vaccine in Adolescents.* N Engl J Med, 2021.
66. Liu, Y., et al., *BNT162b2-Elicited Neutralization against New SARS-CoV-2 Spike Variants.* N Engl J Med, 2021.
67. Callaway, E., *Pfizer COVID vaccine protects against worrying coronavirus variants.* Nature, 2021. **593**(7859): p. 325-326.
68. Abu-Raddad, L.J., et al., *Effectiveness of the BNT162b2 Covid-19 Vaccine against the B.1.1.7 and B.1.351 Variants.* N Engl J Med, 2021.
69. Alaa Abdel Latif, J.L.M., Manar Alkuzweny, Ginger Tsueng, Marco Cano, Emily Haag, Jerry Zhou, Mark Zeller, Nate Matteson, Chunlei Wu, Kristian G. Andersen, Andrew I. Su, Karthik Gangavarapu, Laura D. Hughes, and the Center for Viral Systems Biology, *Lineage Comparison*. 2021: outbreak.info.
70. (WHO), W.H.O. *Tracking SARS-CoV-2 variants*. 2021; Available from: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/.
71. (WHO), W.H.O. *WHO announces simple, easy-to-say labels for SARS-CoV-2 Variants of Interest and Concern*. 2021 31 May; Available from: https://www.who.int/news/item/31-05-2021-who-announces-simple-easy-to-say-labels-for-sars-cov-2-variants-of-interest-and-concern.
72. Korber, B., et al., *Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus.* Cell, 2020. **182**(4): p. 812-827 e19.
73. Pereira, F., *SARS-CoV-2 variants combining spike mutations and the absence of ORF8 may be more transmissible and require close monitoring.* Biochem Biophys Res Commun, 2021. **550**: p. 8-14.
74. Weissman, D., et al., *D614G Spike Mutation Increases SARS CoV-2 Susceptibility to Neutralization.* Cell Host Microbe, 2021. **29**(1): p. 23-31 e4.
75. Hou, Y.J., et al., *SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo.* Science, 2020. **370**(6523): p. 1464-1468.
76. Zhang, L., et al., *SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity.* Nat Commun, 2020. **11**(1): p. 6013.
77. Mansbach, R.A., et al., *The SARS-CoV-2 Spike variant D614G favors an open conformational state.* Sci Adv, 2021. **7**(16).

78. Liu, H., et al., *The basis of a more contagious 501Y.V1 variant of SARS-CoV-2.* Cell Res, 2021. **31**(6): p. 720-722.
79. Sabino, E.C., et al., *Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence.* Lancet, 2021. **397**(10273): p. 452-455.
80. Kurtzman, L. *COVID-19 Variant First Found in Other Countries and States Now Seen More Frequently in California*. 2021 19th January; Available from: https://www.ucsf.edu/news/2021/01/419656/covid-19-variant-first-found-other-countries-and-states-now-seen-more.
81. Min, L. and Q. Sun, *Antibodies and Vaccines Target RBD of SARS-CoV-2.* Front Mol Biosci, 2021. **8**: p. 671633.
82. Jungreis, I., R. Sealfon, and M. Kellis, *SARS-CoV-2 gene content and COVID-19 mutation impact by comparing 44 Sarbecovirus genomes.* Nat Commun, 2021. **12**(1): p. 2642.
83. Hassan, S.S., et al., *A unique view of SARS-CoV-2 through the lens of ORF8 protein.* Comput Biol Med, 2021. **133**: p. 104380.
84. Velazquez-Salinas, L., et al., *Positive Selection of ORF1ab, ORF3a, and ORF8 Genes Drives the Early Evolutionary Trends of SARS-CoV-2 During the 2020 COVID-19 Pandemic.* Front Microbiol, 2020. **11**: p. 550674.
85. Meinberger, D., et al., *Analysis of IgM, IgA, and IgG isotype antibodies Directed against SARS-CoV-2 spike glycoprotein and ORF8 in the course of COVID-19.* Sci Rep, 2021. **11**(1): p. 8920.
86. Wang, X., et al., *Accurate Diagnosis of COVID-19 by a Novel Immunogenic Secreted SARS-CoV-2 orf8 Protein.* mBio, 2020. **11**(5).
87. Pereira, F., *SARS-CoV-2 variants lacking a functional ORF8 may reduce accuracy of serological testing.* J Immunol Methods, 2021. **488**: p. 112906.
88. Flower, T.G., et al., *Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein.* Proc Natl Acad Sci U S A, 2021. **118**(2).
89. Chinese, S.M.E.C., *Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China.* Science, 2004. **303**(5664): p. 1666-9.
90. Muth, D., et al., *Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission.* Sci Rep, 2018. **8**(1): p. 15177.
91. Can, H., et al., *In silico discovery of antigenic proteins and epitopes of SARS-CoV-2 for the development of a vaccine or a diagnostic approach for COVID-19.* Sci Rep, 2020. **10**(1): p. 22387.

92. Su, Y.C.F., et al., *Discovery and Genomic Characterization of a 382-Nucleotide Deletion in ORF7b and ORF8 during the Early Evolution of SARS-CoV-2.* mBio, 2020. **11**(4).
93. Young, B.E., et al., *Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study.* Lancet, 2020. **396**(10251): p. 603-611.
94. Alkhansa, A., G. Lakkis, and L. El Zein, *Mutational analysis of SARS-CoV-2 ORF8 during six months of COVID-19 pandemic.* Gene Rep, 2021. **23**: p. 101024.
95. Stukalov, A., et al., *Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV.* Nature, 2021.
96. Chasapis, C.T., et al., *A SARS-CoV-2 -human metalloproteome interaction map.* J Inorg Biochem, 2021. **219**: p. 111423.
97. Zhang, Y., et al., *The ORF8 protein of SARS-CoV-2 mediates immune evasion through down-regulating MHC-Iota.* Proc Natl Acad Sci U S A, 2021. **118**(23).
98. Li, J.Y., et al., *The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway.* Virus Res, 2020. **286**: p. 198074.
99. Lin, X., et al., *ORF8 contributes to cytokine storm during SARS-CoV-2 infection by activating IL-17 pathway.* iScience, 2021. **24**(4): p. 102293.
100. Wright, J.F., et al., *The human IL-17F/IL-17A heterodimeric cytokine signals through the IL-17RA/IL-17RC receptor complex.* J Immunol, 2008. **181**(4): p. 2799-805.
101. Gaffen, S.L., *Structure and signalling in the IL-17 receptor family.* Nat Rev Immunol, 2009. **9**(8): p. 556-67.
102. Wu, B., et al., *Evolution of the IL17 receptor family in chordates: a new subfamily IL17REL.* Immunogenetics, 2011. **63**(12): p. 835-45.
103. Sohda, M., et al., *Identification of a soluble isoform of human IL-17RA generated by alternative splicing.* Cytokine, 2013. **64**(3): p. 642-5.
104. Zaretsky, M., et al., *Directed evolution of a soluble human IL-17A receptor for the inhibition of psoriasis plaque formation in a mouse model.* Chem Biol, 2013. **20**(2): p. 202-11.
105. Brembilla, N.C., L. Senra, and W.H. Boehncke, *The IL-17 Family of Cytokines in Psoriasis: IL-17A and Beyond.* Front Immunol, 2018. **9**: p. 1682.
106. Yu, J.J. and S.L. Gaffen, *Interleukin-17: a novel inflammatory cytokine that bridges innate and adaptive immunity.* Front Biosci, 2008. **13**: p. 170-7.
107. Gurczynski, S.J. and B.B. Moore, *IL-17 in the lung: the good, the bad, and the ugly.* Am J Physiol Lung Cell Mol Physiol, 2018. **314**(1): p. L6-L16.

108. Foulkes, A.C. and R.B. Warren, *Brodalumab in psoriasis: evidence to date and clinical potential.* Drugs Context, 2019. **8**: p. 212570.
109. Shan, Y., et al., *Preclinical development of GR1501, a human monoclonal antibody that neutralizes interleukin-17A.* Biochem Biophys Res Commun, 2019. **517**(2): p. 303-309.
110. Azevedo, M.L.V., et al., *The role of IL-17A/IL-17RA and lung injuries in children with lethal non-pandemic acute viral pneumonia.* Immunobiology, 2020. **225**(4): p. 151981.
111. Mengesha, B.G. and H.R. Conti, *The Role of IL-17 in Protection against Mucosal Candida Infections.* J Fungi (Basel), 2017. **3**(4).
112. Isailovic, N., et al., *Interleukin-17 and innate immunity in infections and chronic inflammation.* J Autoimmun, 2015. **60**: p. 1-11.
113. Levy, R., et al., *Genetic, immunological, and clinical features of patients with bacterial and fungal infections due to inherited IL-17RA deficiency.* Proc Natl Acad Sci U S A, 2016. **113**(51): p. E8277-E8285.
114. Molesworth-Kenyon, S.J., et al., *IL-17 receptor signaling influences virus-induced corneal inflammation.* J Leukoc Biol, 2008. **83**(2): p. 401-8.
115. Keshavarz, M., et al., *Association of polymorphisms in inflammatory cytokines encoding genes with severe cases of influenza A/H1N1 and B in an Iranian population.* Virol J, 2019. **16**(1): p. 79.
116. Mukherjee, S., et al., *IL-17-induced pulmonary pathogenesis during respiratory viral infection and exacerbation of allergic disease.* Am J Pathol, 2011. **179**(1): p. 248-58.
117. Raffatellu, M., et al., *Simian immunodeficiency virus-induced mucosal interleukin-17 deficiency promotes Salmonella dissemination from the gut.* Nat Med, 2008. **14**(4): p. 421-8.
118. Robinson, K.M., et al., *Influenza A exacerbates Staphylococcus aureus pneumonia by attenuating IL-1beta production in mice.* J Immunol, 2013. **191**(10): p. 5153-9.
119. Guabiraba, R., et al., *IL-22 modulates IL-17A production and controls inflammation and tissue damage in experimental dengue infection.* Eur J Immunol, 2013. **43**(6): p. 1529-44.
120. Neupane, B., et al., *Interleukin-17A Facilitates Chikungunya Virus Infection by Inhibiting IFN-alpha2 Expression.* Front Immunol, 2020. **11**: p. 588382.
121. Leisman, D.E., et al., *Cytokine elevation in severe and critical COVID-19: a rapid systematic review, meta-analysis, and comparison with other inflammatory syndromes.* Lancet Respir Med, 2020. **8**(12): p. 1233-1244.
122. Rubin, E.J., D.L. Longo, and L.R. Baden, *Interleukin-6 Receptor Inhibition in Covid-19 - Cooling the Inflammatory Soup.* N Engl J Med, 2021. **384**(16): p. 1564-1565.

123. Fricke-Galindo, I. and R. Falfan-Valencia, *Genetics Insight for COVID-19 Susceptibility and Severity: A Review.* Front Immunol, 2021. **12**: p. 622176.
124. Gordon, D.E., et al., *Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms.* Science, 2020. **370**(6521).
125. Cacciapuoti, S., et al., *Immunocytometric analysis of COVID patients: A contribution to personalized therapy?* Life Sci, 2020. **261**: p. 118355.
126. Curtis, M.M. and S.S. Way, *Interleukin-17 in host defence against bacterial, mycobacterial and fungal pathogens.* Immunology, 2009. **126**(2): p. 177-85.
127. Garcia-Vidal, C., et al., *Incidence of co-infections and superinfections in hospitalized patients with COVID-19: a retrospective cohort study.* Clin Microbiol Infect, 2021. **27**(1): p. 83-88.
128. Lv, L., et al., *Gut mycobiota alterations in patients with COVID-19 and H1N1 infections and their associations with clinical features.* Commun Biol, 2021. **4**(1): p. 480.
129. Hoenigl, M., *Invasive Fungal Disease complicating COVID-19: when it rains it pours.* Clin Infect Dis, 2020.
130. Gandotra, S. and D. Russell, *The Long and the Short of It: Is "Long COVID" More Than Slow Resolution of the Acute Disease?* Ann Am Thorac Soc, 2021. **18**(6): p. 948-950.
131. Stengel, A., et al., *Long Haulers-What Is the Evidence for Post-COVID Fatigue?* Front Psychiatry, 2021. **12**: p. 677934.
132. Sivan, M., C. Rayner, and B. Delaney, *Fresh evidence of the scale and scope of long covid.* BMJ, 2021. **373**: p. n853.
133. Ayoubkhani, D., et al., *Post-covid syndrome in individuals admitted to hospital with covid-19: retrospective cohort study.* BMJ, 2021. **372**: p. n693.
134. Pacha, O., M.A. Sallman, and S.E. Evans, *COVID-19: a case for inhibiting IL-17?* Nat Rev Immunol, 2020. **20**(6): p. 345-346.
135. Orlov, M., et al., *A Case for Targeting Th17 Cells and IL-17A in SARS-CoV-2 Infections.* J Immunol, 2020. **205**(4): p. 892-898.
136. Berman, J.J., *Group IV Viruses: Single-Stranded (+)Sense RNA.* Taxonomic Guide to Infectious Diseases, 2012: p. 237-246.
137. Onyango, M.G., A.T. Ciota, and L.D. Kramer, *The Vector - Host - Pathogen Interface: The Next Frontier in the Battle Against Mosquito-Borne Viral Diseases?* Front Cell Infect Microbiol, 2020. **10**: p. 564518.

138. Shah, P.S., et al., *Comparative Flavivirus-Host Protein Interaction Mapping Reveals Mechanisms of Dengue and Zika Virus Pathogenesis.* Cell, 2018. **175**(7): p. 1931-1945 e18.
139. Beaver, J.T., et al., *Evolution of Two Major Zika Virus Lineages: Implications for Pathology, Immune Response, and Vaccine Development.* Front Immunol, 2018. **9**: p. 1640.
140. Guo, M., et al., *ZIKV viral proteins and their roles in virus-host interactions.* Sci China Life Sci, 2021. **64**(5): p. 709-719.
141. Estevez-Herrera, J., et al., *Zika Virus Pathogenesis: A Battle for Immune Evasion.* Vaccines (Basel), 2021. **9**(3).
142. Yun, S.I., et al., *Complete Genome Sequences of Three Historically Important, Spatiotemporally Distinct, and Genetically Divergent Strains of Zika Virus: MR-766, P6-740, and PRVABC-59.* Genome Announc, 2016. **4**(4).
143. Kuno, G. and G.J. Chang, *Biological transmission of arboviruses: reexamination of and new insights into components, mechanisms, and unique traits as well as their evolutionary trends.* Clin Microbiol Rev, 2005. **18**(4): p. 608-37.
144. Tan, T.Y., et al., *Capsid protein structure in Zika virus reveals the flavivirus assembly process.* Nat Commun, 2020. **11**(1): p. 895.
145. Barnard, T.R., et al., *Molecular Determinants of Flavivirus Virion Assembly.* Trends Biochem Sci, 2021. **46**(5): p. 378-390.
146. Tan, T.Y., G. Fibriansah, and S.M. Lok, *Capsid protein is central to the birth of flavivirus particles.* PLoS Pathog, 2020. **16**(5): p. e1008542.
147. Kostyuchenko, V.A., et al., *Immature and mature dengue serotype 1 virus structures provide insight into the maturation process.* J Virol, 2013. **87**(13): p. 7700-7.
148. Li, T., et al., *Structural insight into the Zika virus capsid encapsulating the viral genome.* Cell Res, 2018. **28**(4): p. 497-499.
149. Oliveira, E.R.A., et al., *The flavivirus capsid protein: Structure, function and perspectives towards drug design.* Virus Res, 2017. **227**: p. 115-123.
150. Prasad, V.M., et al., *Structure of the immature Zika virus at 9 A resolution.* Nat Struct Mol Biol, 2017. **24**(2): p. 184-186.
151. Sirohi, D., et al., *The 3.8 A resolution cryo-EM structure of Zika virus.* Science, 2016. **352**(6284): p. 467-70.
152. Shang, Z., et al., *Crystal Structure of the Capsid Protein from Zika Virus.* J Mol Biol, 2018. **430**(7): p. 948-962.

153. Scaturro, P., A.L. Kastner, and A. Pichlmair, *Chasing Intracellular Zika Virus Using Proteomics.* Viruses, 2019. **11**(9).
154. Scaturro, P., et al., *An orthogonal proteomic survey uncovers novel Zika virus host factors.* Nature, 2018. **561**(7722): p. 253-257.
155. Fontaine, K.A., et al., *The Cellular NMD Pathway Restricts Zika Virus Infection and Is Targeted by the Viral Capsid Protein.* mBio, 2018. **9**(6).
156. Gestuveo, R.J., et al., *Analysis of Zika virus capsid-Aedes aegypti mosquito interactome reveals pro-viral host factors critical for establishing infection.* Nat Commun, 2021. **12**(1): p. 2766.
157. Cloherty, A.P.M., et al., *Hijacking of Lipid Droplets by Hepatitis C, Dengue and Zika Viruses-From Viral Protein Moonlighting to Extracellular Release.* Int J Mol Sci, 2020. **21**(21).
158. Hou, S., et al., *Zika Virus Hijacks Stress Granule Proteins and Modulates the Host Stress Response.* J Virol, 2017. **91**(16).
159. Li, G., et al., *Characterization of cytopathic factors through genome-wide analysis of the Zika viral proteins in fission yeast.* Proc Natl Acad Sci U S A, 2017. **114**(3): p. E376-E385.
160. Zeng, J., et al., *The Zika Virus Capsid Disrupts Corticogenesis by Suppressing Dicer Activity and miRNA Biogenesis.* Cell Stem Cell, 2020. **27**(4): p. 618-632 e9.
161. Slomnicki, L.P., et al., *Ribosomal stress and Tp53-mediated neuronal apoptosis in response to capsid protein of the Zika virus.* Sci Rep, 2017. **7**(1): p. 16652.
162. Leon, K., et al., *Zika Virus Infection Prevents Host mRNA Nuclear Export by Disrupting UPF1 Function.* bioRxiv, 2020: p. 2020.12.03.410837.
163. Leon, K. and M. Ott, *An 'Arms Race' between the Nonsense-mediated mRNA Decay Pathway and Viral Infections.* Semin Cell Dev Biol, 2021. **111**: p. 101-107.
164. Li, M., et al., *Identification of antiviral roles for the exon-junction complex and nonsense-mediated decay in flaviviral infection.* Nat Microbiol, 2019. **4**(6): p. 985-995.
165. Ramage, H.R., et al., *A combined proteomics/genomics approach links hepatitis C virus infection with nonsense-mediated mRNA decay.* Mol Cell, 2015. **57**(2): p. 329-340.
166. Jaffrey, S.R. and M.F. Wilkinson, *Nonsense-mediated RNA decay in the brain: emerging modulator of neural development and disease.* Nat Rev Neurosci, 2018. **19**(12): p. 715-728.
167. Vicente-Crespo, M. and I.M. Palacios, *Nonsense-mediated mRNA decay and development: shoot the messenger to survive?* Biochem Soc Trans, 2010. **38**(6): p. 1500-5.

168. Rosa-Fernandes, L., et al., *Zika Virus Impairs Neurogenesis and Synaptogenesis Pathways in Human Neural Stem Cells and Neurons.* Front Cell Neurosci, 2019. **13**: p. 64.
169. Hubel, P., et al., *A protein-interaction network of interferon-stimulated genes extends the innate immune system landscape.* Nat Immunol, 2019. **20**(4): p. 493-502.
170. Luck, K., et al., *A reference map of the human binary protein interactome.* Nature, 2020. **580**(7803): p. 402-408.
171. Bojkova, D., et al., *Proteomics of SARS-CoV-2-infected host cells reveals therapy targets.* Nature, 2020. **583**(7816): p. 469-472.
172. Hafirassou, M.L., et al., *A Global Interactome Map of the Dengue Virus NS1 Identifies Virus Restriction and Dependency Host Factors.* Cell Rep, 2017. **21**(13): p. 3900-3913.

Chapter 2

Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant

Xianding Deng,# Miguel A. Garcia-Knight,# **Mir M. Khalid**,# Venice Servellita,# Candace Wang,# Mary Kate Morris,# Alicia Sotomayor-Gonza´ lez, Dustin R. Glasner, Kevin R. Reyes, Amelia S. Gliwa, Nikitha P. Reddy, Claudia Sanchez San Martin, Scot Federman, Jing Cheng, Joanna Balcerek, Jordan Taylor, Jessica A. Streithorst, Steve Miller, Bharath Sreekumar, Pei-Yi Chen, Ursula Schulze-Gahmen, Taha Y. Taha, Jennifer M. Hayashi, Camille R. Simoneau, G. Renuka Kumar, Sarah McMahon, Peter V. Lidsky, Yinghong Xiao, Peera Hemarajata, Nicole M. Green, Alex Espinosa, Chantha Kath, Monica Haw, John Bell, Jill K. Hacker, Carl Hanson, Debra A. Wadford, Carlos Anaya, Donna Ferguson, Phillip A. Frankino, Haridha Shivram, Liana F. Lareau, Stacia K. Wyman, Melanie Ott,* Raul Andino,* and Charles Y. Chiu.*

Cell 184, 1–12(2021)

DOI:10.1016/j.cell.2021.04.025

#These authors contributed equally *Lead contact

Transmission, infectivity, and antibody neutralization of an emerging SARS-CoV-2 variant in California carrying a L452R spike protein mutation

Xianding Deng1,2&, Miguel A Garcia-Knight3&, Mir M. Khalid4,5&, Venice Servellita1,2&, Candace Wang1,2&, Mary Kate Morris6&, Alicia Sotomayor-González1,2, Dustin R Glasner1,2, Kevin R Reyes1,2, Amelia S. Gliwa1,2, Nikitha P. Reddy1,2, Claudia Sanchez San Martin1,2, Scot Federman7, Jing Cheng4, Joanna Balcerek1, Jordan Taylor1, Jessica A Streithorst1, Steve Miller1, G. Renuka Kumar4,5, Bharath Sreekumar4,5, Pei-Yi Chen4,5, Ursula Schulze-Gahmen4,5, Taha Y. Taha4,5 , Jennifer Hayashi4,5, Camille R. Simoneau4,5, Sarah McMahon4,5, Peter V. Lidsky3, Yinghong Xiao3, Peera Hemarajata8, Nicole M. Green8, Alex Espinosa6, Chantha Kath6, Monica Haw6, John Bell6, Jill K. Hacker6, Carl Hanson6, Debra A. Wadford6, Carlos Anaya9, Donna Ferguson9, Liana F. Lareau10,11, Phillip A. Frankino11, Haridha Shivram11, Stacia K. Wyman11, Melanie Ott4,5,11*, Raul Andino3*, Charles Y. Chiu1,2,4,11*#

1Department of Laboratory Medicine, University of California San Francisco, California, USA

2UCSF-Abbott Viral Diagnostics and Discovery Center, San Francisco, California, USA

3Department of Microbiology and Immunology, University of California San Francisco, California, USA

4Department of Medicine, University of California San Francisco, California, USA

5Gladstone Institute of Virology, San Francisco, California, USA

6California Department of Public Health, Richmond, California, USA

7Laboratory for Genomics Research, University of California San Francisco, California, USA

8Los Angeles County Department of Public Health, Los Angeles, California, USA

9Monterey County Department of Public Health, Monterey, California, USA

10Department of Bioengineering, University of California Berkeley, Berkeley, California, USA

11Innovative Genomics Institute, University of California Berkeley, Berkeley, California, USA

&co-first authors

*co-senior authors

#corresponding author: Charles Y. Chiu, 185 Berry Street, Box #0134, San Francisco, CA 94107, USA; e-mail: charles.chiu@ucsf.edu

Abstract

We identified an emerging severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variant by viral whole-genome sequencing of 2,172 nasal/nasopharyngeal swab samples from 44 counties in California, a state in the western United States. Named B.1.427/B.1.429 to denote its two lineages, the variant emerged in May 2020 and increased from 0% to >50% of sequenced cases from September 2020 to January 2021, showing 18.6%–24% increased transmissibility relative to wild-type circulating strains. The variant carries three mutations in the spike protein, including an L452R substitution. We found 2-fold increased B.1.427/B.1.429 viral shedding in vivo and increased L452R pseudovirus infection of cell cultures and lung organoids, albeit decreased relative to pseudoviruses carrying the N501Y mutation common to variants B.1.1.7, B.1.351, and P.1. Antibody neutralization assays revealed 4.0- to 6.7-fold and 2.0-fold decreases in neutralizing titers from convalescent patients and vaccine recipients, respectively. The increased prevalence of a more transmissible variant in California exhibiting decreased antibody neutralization warrants further investigation.

Key words: SARS-CoV-2; coronavirus; pandemic; COVID-19; viral whole-genome sequencing; genomic surveillance; molecular dating; genomic epidemiology; spike protein; L452R mutation; variant; antibody neutralization; vaccine; N501Y mutation; B.1.427/B.1.429; 20C/L452R; pseudovirus infectivity studies; antibody neutralization

Introduction

Genetic mutation provides a mechanism for viruses to adapt to a new host and/or evade host immune responses. Although severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has a slow evolutionary rate relative to other RNA viruses (~0.8 × 10−3 substitutions per site per year) (Day et al., 2020), an unabating coronavirus disease 2019 (COVID-19) pandemic with high viral transmission has enabled the virus to acquire significant genetic diversity since its initial detection in Wuhan, China in December 2019 (Zhu et al., 2020), thereby facilitating the emergence of new variants (Fontanet et al., 2021). Among numerous SARS-CoV-2 variants now circulating globally, those harboring a D614G mutation have predominated since June of 2020 (Korber et al., 2020), possibly due to enhanced viral fitness and transmissibility (Hou et al., 2020; Plante et al., 2021; Zhou et al., 2021).

Emerging variants of SARS-CoV-2 that harbor genome mutations that may impact transmission, virulence, and immunity have been designated “variants of concern” (VOCs). Beginning in the fall of 2020, 3 VOCs have emerged globally, each carrying multiple mutations across the genome, including several in the receptor-binding domain (RBD) of the spike protein. The B.1.1.7 variant, originally detected in the United Kingdom (UK) (Chand et al., 2020), has accumulated 17 lineage-defining mutations, including the spike protein N501Y mutation that confers increased transmissibility over other circulating viruses (Leung et al., 2021; Rambaut et al., 2020b; Volz et al., 2020). Preliminary data suggest that B.1.1.7 may also cause more severe illness (Davies et al., 2021b). As of early 2021, the B.1.1.7 variant has become the predominant lineage throughout the United Kingdom and Europe, with reported cases also rising in the United States (US) (Washington et al., 2021). The other two VOCs, B.1.351 detected in South Africa (Tegally et al., 2020) and P.1 first detected in Brazil (Sabino et al., 2021), carry E484K and K417N/K417T in addition to N501Y mutations. Multiple studies have reported that the E484K mutation in particular may confer resistance to antibody neutralization (Cole et al., 2021; Wang et al., 2021; Wibmer et al., 2021; Wu et al., 2021b; Xie et al., 2021), potentially resulting in decreased effectiveness of currently available vaccines (Liu et al., 2021; Wise, 2021). This phenotype may have also contributed to widespread reinfection by P.1 in an Amazon community that had presumptively achieved herd immunity (Buss et al., 2021; Sabino et al., 2021).

In January 2021, we and others independently reported the emergence of a variant in California carrying an L452R mutation in the RBD of the spike protein (CDPH (California Department of Public Health), 2021a; Zhang et al., 2021). Here, we used viral whole-genome sequencing of nasal/nasopharyngeal (N/NP) swab samples from multiple counties to characterize the

emergence and spread of this L452R-carrying variant in California from September 1, 2020, to January 29, 2021. We also combined epidemiologic, clinical, and in vitro laboratory data to investigate transmissibility and susceptibility to antibody neutralization associated with infection by the variant.

Results

Viral genomic surveillance

We sequenced 2,172 viral genomes across 44 California counties from remnant N/NP swab samples testing positive for SARS-CoV-2 (Tables S1 and S2). The counties with proportionally higher representation in the dataset included Santa Clara County (n = 725, 33.4%), Alameda County (n = 228, 10.5%), Los Angeles County (n = 168, 7.7%), and San Francisco County (n = 155, 7.1%) (Figure 1A). A variant, subsequently named 20C/L452R according to the Nextstrain nomenclature system (Bedford et al., 2021) or B.1.427/B.1.429 according to the Pango system (Rambaut et al., 2020a) (henceforth referred to using the Pango designation to distinguish between the B.1.427 and B.1.429 lineages), was identified in 21.5% (466 of 2,172) of the genomes (Table S1). The frequency of this variant in California increased from 0% at the beginning of September 2020 to >50% of sequenced cases by the end of January 2021, with a similar trajectory to the surge of COVID-19 cases in California from October to December 2020 (Figure S1). The rise in the proportion of sequenced cases due to the variant was rapid, with an estimated increase in transmission rate of the B.1.427/B.1.429 variant relative to circulating non-B.1.427/B.1.429 lineages of 20.0% (17.8%–21.1%) and an approximate doubling time of 19.1 days (17.3–21.4 days) (Figure 1B, top panel). The calculated date for when the variant was expected to become predominant (>50% of cases) in California was January 25, 2021, earlier but near in time to the February 5, 2021 date based on additional viral genomic data from samples collected February 1–March 11, 2021 (Figure 1B, bottom panel). Similar epidemic trajectories were observed from multiple counties (Figures 1C–1E and S2), despite different sampling approaches used for sequencing. Specifically, genomes from San Francisco County were derived from COVID-19 patients being tested at University of California, San Francisco (UCSF) hospitals and clinics; genomes from Alameda County were derived from community testing; genomes from Santa Clara County were derived from congregate facility, community, and acute care testing; and genomes from Los Angeles County were derived from coroner, community, and inpatient testing.

Phylogenetic and molecular dating analyses

Bayesian phylogenetic analysis of 1,153 genomes subsampled from a 2,519-genome dataset consisting of the 2,172 California genomes sequenced in this study and 347 representative global genomes (Bedford and Neher, 2020) identified two distinct lineages in clade 20C (Nextstrain designation) associated with the variant, B.1.427 and B.1.429 (Figures 2A and 2C). Both lineages share a triad of coding mutations in the spike protein (S13I, W152C, and L452R), one coding mutation in the orf1b protein (D1183Y), and an additional 2 non-coding mutations (Figure 2A). Four additional mutations, one of them a coding mutation in orf1a (I4205V), were specific to B.1.429, while 3 additional mutations, including two coding mutations in orf1a (S3158T) and orf1b (P976L), were specific to B.1.427. A root-to-tip genetic distance plot of the 1,153 subsampled genomes showed no substantial difference between B.1.427/B.1.429 variant and non-variant lineages (Figure 2B).

Using a previously reported algorithm to assess divergence time dating (Drummond et al., 2012), we estimated that the most recent common ancestor emerged on May 4, 2020 (95% highest posterior density [HPD] interval: April 10–May 29). The branches giving rise to the B.1.429 and B.1.427 lineages were predicted to have diverged on May 30 (95% HPD: May 10–June 18) and August 7 (95% HPD: June 22–September 18), respectively (Figure 2D).

Increased transmissibility and infectivity

Analysis of available data from 2,126 (97.8%) of the 2,172 sequenced genomes in the current study revealed that the median PCR cycle threshold (Ct) value associated with B.1.427/B.1.429 variant infections was significantly lower overall ($p = 4.75 \times 10{-}7$) than that associated with non-variant viruses (Figure 3A). We estimated that in swab samples N/NP viral RNA loads are ~2-fold higher in B.1.427/B.1.429 than in non-variant viruses (Drew et al., 2020). The differences in cycle threshold were greater during the November and December months relative to January (Figure 3B), although these differences were not statistically significant due to lower sample numbers. There did not appear to be significant differences in cycle threshold between hospitalized patients and outpatients infected with B.1.427/B.1.429 (Figure 3C), nor between B.1.427 and B.1.429 lineages (Figure 3D).

Analysis of the SARS-CoV-2 spike protein complexed to its human ACE2 receptor (Lan et al., 2020) revealed that the L452 residue does not directly contact the receptor. Instead, L452, together with F490 and L492, form a hydrophobic patch on the surface of the spike RBD (Figure 4A). To understand the effects of L452R RBD mutation on viral entry, pseudoviruses carrying

D614G with L452R or W152C, or D614G alone were generated and used for infection of 293T cells stably expressing the ACE2 cell entry receptor and TMPRSS2 cofactor for SARS-CoV-2 (Hoffmann et al., 2020; Khanna et al., 2020). We observed increased entry by pseudoviruses carrying the L452R mutation compared to D614G alone, with a 6.7- to 22.5-fold increase in 293T cells and a 5.8- to 14.7-fold increase in HAOs (Figures 4B and 4C). This increase in pseudovirus infection with the L452R mutation is slightly lower than the increase observed with the N501Y mutation (11.4- to 30.9-fold increase in 293T cells and 23.5- to 37.8-fold increase in HAO relative to D614G alone), which has previously been reported to increase pseudovirus entry (Hu et al., 2021). Pseudoviruses carrying the W152C mutation demonstrated small increases in infection of 293T cells and HAO relative to the D614 control, although these increases were not as pronounced as those observed for the L452R and N501Y pseudoviruses.

Reduced susceptibility to neutralizing antibodies from convalescent patients and vaccine recipients

To examine the effect of the L452R mutation on antibody binding, we performed neutralizing antibody assays. We cultured a B.1.429 lineage virus from a patient's NP swab sample in Vero cells stably expressing TMPRSS2 (Vero-TMPRSS2). We then performed plaque reduction neutralization tests (PRNT) using 21 plasma samples from convalescent patients and vaccine recipients to compare neutralization titers between the B.1.429 isolate and a control isolate USA-WA1/2020 (Figures 5A and S3; Table S3). Twelve samples were collected from individuals after receiving both doses of either the Pfizer BNT16b2 or Moderna mRNA-1273 vaccine, with samples collected 4–28 days after the second dose. Nine samples were convalescent plasma collected from patients who became symptomatic from COVID-19 infection during the June 21–November 11, 2020 time period, during which infection from a VOC in California was highly unlikely. Convalescent samples were collected 18–71 days after symptom onset. Measurable neutralizing antibody responses in the assay range were not observed for 1 convalescent patient and 1 vaccine recipient.

We found that in comparison to USA-WA1/2020, 7 of 8 (88%) convalescent patients and 6 of 11 (55%) vaccine recipients showed reduced PRNT50 titers to a B.1.429 lineage virus, with 6.7-fold ($p = 0.016$) and 2-fold ($p = 0.031$) median reductions, respectively (Figure 5A). There were no differences in neutralization between WA1 and D614G isolates by convalescent or post-vaccination plasma (Figure 5A, right).

Next, we independently evaluated neutralizing antibody responses against a cultured B.1.427 lineage virus. The TCID50, or median tissue culture infective dose at which 50% of cultures exhibited cytopathic effect (CPE), was determined for 10 different convalescent plasma samples collected from COVID-19 patients from June 19 to August 19, 2020, with samples collected 21–85 days after symptom onset. Nine of 10 (90%) convalescent patients showed reduced TCID50 titers to a B.1.427 lineage virus (Figure 5B), with 5.3-fold ($p = 0.0039$) and 4.0-fold ($p = 0.0039$) median reductions compared to USA-WA1/2020) and D614G isolates, respectively.

Discussion

As of early 2021, multiple SARS-CoV-2 variants have emerged in different regions of the world, each rapidly establishing itself as the predominant lineage within a few months after its initial detection (Chand et al., 2020; Faria et al., 2021; Sabino et al., 2021; Tegally et al., 2020). In the current study, we describe the spread of an emerging B.1.427/B.1.429 variant in California carrying a characteristic triad of spike protein mutations (S13I, W152C, and L452R) that is predicted to have emerged in May 2020 and increased in frequency from 0% to >50% of sequenced cases from September 2020 to January 2021. Importantly, this variant was found to comprise 2 separate lineages, B.1.427 and B.1.429, with each lineage rising in parallel in California as well as in multiple other states (Gangavarapu et al., 2020). Potential increased transmissibility of the B.1.427/B.1.429 variant is also supported by findings of an ~2-fold increase in median viral loads in infected patients and increased infectivity of cultured cells and lung organoids in vitro. We also observed a moderate resistance to neutralization by antibodies elicited by prior infection (4.0- to 6.7-fold) or vaccination (2-fold). These findings indicate that the B.1.427/B.1.429 variant warrants close monitoring and further investigation regarding its potential to cause future surges in COVID-19 cases, accumulate further mutations, and/or decrease vaccine effectiveness.

The results here highlight the urgent need for implementation of a robust genomic surveillance system in the United States and globally to rapidly identify and monitor SARS-CoV-2 variants. Although our findings suggest that the B.1.427/B.1.429 variant emerged as early as May 2020, the first cases of B.1.427 and B.1.429 in the United States were not identified by sequencing until September 28, 2020, and July 13, 2020, respectively. Sparse genomic sequencing of circulating viruses likely contributed to delayed identification of the B.1.427/B.1.429 variant. Furthermore, unlike in countries such as the United Kingdom (COVID-19 Genomics UK (COG-UK), 2020) and South Africa (Msomi et al., 2020), the United States lacks an organized system for real-time analysis and reporting of variants that is tied to actionable public health responses. Public

disclosure of the existence of this variant, initiated by us in coordination with local and state public health agencies and the United States Centers for Disease Control and Prevention (US CDC), did not occur until January 17, 2021 (CDPH (California Department of Public Health), 2021a), by which time the variant had already become the dominant lineage in several California counties and spread to multiple other states (Gangavarapu et al., 2020). Earlier identification and monitoring of the variant might have guided focused contact tracing efforts by public health to slow its spread, as well as enabled more timely investigation of its potential significance. Our identification of the B.1.427/B.1.429 variant was made possible by California COVIDNet, a collaborative sequencing network working to track transmission and evolution of SARS-CoV-2 in the state by viral whole-genome sequencing (CDPH (California Department of Public Health), 2021a).

The B.1.427/B.1.429 variant carries 4 coding mutations, including 3 in the spike protein, that are not found in the 3 other SARS-CoV-2 VOCs (B.1.1.7, B.1.351, and P.1) or in other major circulating lineages. The appearance of several new mutations in a variant over a short period of time is not unexpected and may be indicative of sudden increase in the evolutionary rate of a directly ancestral lineage (Rambaut et al., 2020b). Indeed, the B.1.1.7, B.1.351, and P.1 variants exhibit striking genetic divergence, with each carrying over 8 missense mutations in the spike protein (Faria et al., 2021; Rambaut et al., 2020b; Tegally et al., 2020). The evolutionary mechanism underlying these changes remains unexplained but may potentially be due to accelerated viral quasispecies evolution in chronically infected patients (Avanzato et al., 2020; Choi et al., 2020; Kemp et al., 2021). In contrast to VOCs such as B.1.1.7 (Rambaut et al., 2020b), the root-to-tip divergence plot corresponding to the B.1.427/B.1.429 variant is consistent with gradual accumulation of mutations over time. However, we also cannot rule out accelerated evolution of the variant given the absence of sequenced genomes directly ancestral to the B.1.427 and B.1.429 lineages, possibly due to limited genomic sampling (Figure S1), as well as the anomalous position of the first sequenced B.1.429 genome from Los Angeles County in July 2020 on the root-to-tip divergence plot.

Prior studies have suggested that the L452R mutation may stabilize the interaction between the spike protein and its human ACE2 receptor and thereby increase infectivity (Chen et al., 2020; Teng et al., 2021). Our findings of enhanced infection of 293T cells and lung organoids by pseudoviruses carrying L452R confirm these early predictions. Notably, the L452 residue does not directly contact the ACE2 receptor, unlike the N501 residue that is mutated to Y501 in the highly transmissible B.1.1.7, B.1.351, and P.1 variants. However, given that L452 is positioned in

a hydrophobic patch of the spike RBD, it is plausible that the L452R mutation causes structural changes in the region that promote the interaction between the spike protein and its ACE2 receptor. Notably, our findings reveal that the infectivity of L452R pseudoviruses was higher than D614G, but slightly reduced compared to that of N501Y pseudoviruses in 293T cells and human airway lung organoids.

Interestingly, we found that the observed differences in viral load were more pronounced during the November and December months, when cases and deaths of COVID-19 in California were surging (CDPH (California Department of Public Health), 2021b), than in December. These findings likely reflect sampling bias, with a possible increased focus on sequencing cases in symptomatic patients and/or associated with outbreaks. Nevertheless, the impact of increased transmissibility associated with B.1.427/B.1.429 on disease severity is a critical question that we are aiming to address in ongoing studies. It is notable that infection by the highly contagious N501Y-carrying B.1.1.7 variant has been shown to be associated with an increased risk of severe disease and death (Challen et al., 2021; Davies et al., 2021c). In addition, whether the L452R-carrying B.1.427/B.1.429 will continue to remain the predominant circulating strain in California, or whether it will eventually be replaced by the B.1.1.7 variant (Washington et al., 2021) remains unclear.

The L452R mutation in the B.1.427/B.1.429 variant has been observed previously in rare, mostly singleton cases, first reported from Denmark on March 17, 2020, and also reported from multiple US states and the United Kingdom prior to September 1, 2020 (Gangavarapu et al., 2020). Given our findings of increased infectivity of L452R pseudoviruses, it is unclear why surges in L452R-carrying lineages have not occurred earlier. We speculate that although these lineages may have been more infective, transmission may not have reached a critical threshold locally or may have been influenced by other factors such as population density and/or public health interventions. An alternative (but not mutually exclusive) possibility is that the additional mutations in B.1.427/B.1.429, especially the W152C and S13I mutations in the spike protein, may contribute to increased infectivity of the variant relative to lineages carrying the L452R mutation alone. Indeed, in the current study we observed smaller but statistically significant increases in infection of 293T cell and lung organoids by pseudoviruses carrying W152C. Studies of pseudoviruses carrying the 3 spike mutations or the full complement of mutations in the B.1.427/B.1.429 variant are needed to address these hypotheses.

Our neutralization findings are consistent with a prior report showing decreased binding of L452R-carrying pseudoviruses by antibodies from previously infected COVID-19 patients and escape

from neutralization in 3 of 4 convalescent plasma samples (Liu et al., 2020). We speculate that mutation of the L452 residue in a hydrophobic pocket may induce conformational changes in the RBD that impact neutralizing antibody binding. Of note, a >4-fold decrease in neutralizing antibody titers in convalescent plasma suggests that immune selection pressure from a previously exposed population may be partly driving the emergence of L452R variants. These data also raise questions regarding potential higher risk of re-infection and the therapeutic effectiveness of monoclonal antibodies and convalescent plasma to treat COVID-19 disease from the B.1.427/B.1.429 variant.

Overall, the modest 2-fold decrease in neutralizing antibody titers in vaccine recipients to the B.1.429 variant is an indication of the robust neutralizing antibody responses elicited by mRNA vaccines in the face of variants under immune selection pressure. Indeed, a reduction in neutralization of a similar magnitude associated with the L452R mutation has been reported following mRNA vaccination in studies using pseudotype assays (Garcia-Beltran et al., 2021; Wu et al., 2021a). The use of a B.1.429 isolate in the present study, carrying the full complement of mutations that characterize the lineage, may account for relative fold differences between these two aforementioned studies and ours, and the contribution of epistatic mutations to neutralization phenotypes for SARS-CoV-2 variants merits further study. In addition, because neutralizing antibodies in natural infection have been shown to wane over time (Lau et al., 2021; Seow et al., 2020), longitudinal serologic studies are needed to determine whether these modest decreases will affect the long-term durability of vaccine-elicited immune responses to the B.1.427/B.1.429 variant. Of concern is also the possibility that B.1.427/B.1.429 lineages may accumulate additional mutations in the future that may further enhance the escape phenotype.

Limitations of study

Although in this study, we obtain robust estimates for the emergence and growth of the B.1.427/B.1.429, these estimates may be biased by uneven sampling and limited genomic sampling overall relative to the number of COVID-19 infections in California (Figure S1). The pseudovirus infectivity studies evaluated only the L452R mutation, and the impact of other mutations in the B.1.427/B.1.429 genome in combination needs to be studied experimentally. The neutralization studies included a limited number of convalescent patients (n = 19 in total) and vaccine recipients (n = 12); in addition, some of the vaccine recipients had not yet received the second dose or were sampled prior to 14 days after their second dose. Further investigation of potential antibody neutralization escape associated with the B.1.427/B.1.429 variant in larger cohorts of patients and vaccinees is needed to confirm our results.

Figure 1. IIncreasing frequency of the B.1.427/B.1.429 variant in California from September 1, 2020 to January 29, 2021 .

(A) County-level representation of the 2,172 newly sequenced SARS-CoV-2 genomes in the current study. Counties from which at least 1 genome were sequenced are colored in sky blue. The size of the circle is proportionate to the number of genomes sequenced from each county, while points designate counties where fewer than 10 genomes were sequenced.

(B–D) Logistic growth curves fitting the 5-day rolling average of the estimated proportion of B.1.427/B.1.429 variant cases in **(B)** California, **(C)** San Francisco County, and **(D)** Santa Clara County. For each curve, the estimated increase in transmission rate and doubling time are shown, along with their associated 95% confidence intervals. The predicted time when the growth curve crosses 0.5 is indicated by a vertical red line. A vertical black dotted line denotes the transition from 2020 to 2021. **(B) Top**: the logistic growth curve generated from all 2,172 genomes in the current study. The 95% confidence intervals for the increase in transmission rate and doubling time are shaded in blue and gray, respectively. **(B) Bottom**: the logistic growth curve with inclusion of an additional 2,737 sequenced genomes from California collected February 1 to March 11, 2021. The increase in transmission rate is defined as the logistic growth rate multiplied by the serial interval (Volz et al., 2020; Washington et al., 2021).

Figure 2. Genomic, phylogenetic, and molecular clock analyses of the B.1.427/B.1.429 variant in California.

(A) A multiple sequence alignment of 6 representative B.1.427/B.1.429 genomes, 3 from the B.1.427 lineage, and 3 from the B.1.429 lineage, using the prototypical Wuhan Hu-1 genome as a reference. Defining single nucleotide polymorphisms (SNPs) in the B.1.427 and B.1.429 lineages are compared to each other and to other SARS-CoV-2 viruses in Nextstrain clade 20C. The SNPs are color coded as follows: red SNPs are shared between the B.1.427 and B.1.429 lineages, blue SNPs are specific to B.1.427, purple SNPs are specific to B.1.429, brown SNPs are shared with other clade 20C viruses, and gray SNPs are specific to individual viruses.

(B) Root-to-tip divergence plot of number of accumulated mutations by month based on 1,153 genomes subsampled from a complete dataset consisting of the 2,172 genomes recovered in the current study and 347 representative global genomes. The gray highlighted region encompasses the period of sampling for nearly all genomes sequenced in the current study (September 1, 2020 to January 31, 2021), with the exception of the first 2 sequenced B.1.429 genomes from Los Angeles that were reported on July 20, 2020. The orange-red bullseye denotes the first reported

genomic sequence of the B.1.429 variant from Los Angeles County from a sample collected July 13, 2020.

(C) Maximum likelihood circular phylogenetic tree of the 1,153 subsampled genomes, denoting the major viral clades. The red asterisk denotes a UK B.1.1.7 variant genome.

(D) Time scaled maximum clade credibility (MCC) tree, showing the median divergence dates and associated 95% highest posterior density (HPD) distributions, or confidence intervals, for the B.1.427/B.1.429 variant (D1), B.1.429 lineage (D2), and B.1.427 lineage (D3), as estimated from TMRCA (time to most recent common ancestor) calculations. The B.1.427 lineage is colored in blue and the B.1.429 lineage in red. The orange-red bullseye denotes the first reported genomic sequence of the B.1.429 variant from Los Angeles County from a sample collected July 13, 2020.

Figure 3. Higher viral loads in infections from the B.1.427/B.1.429 variant as compared to non-B.1.427/B.1.429 variant lineages.

(A–D) Boxen plots of available PCR cycle threshold (Ct) values for B.1.427/B.1.429 variant compared to non-variant identification for **(A)** all samples sequenced in the current study, **(B)** samples stratified by month of collection, November 2020–January 2021, **(C)** samples from hospitalized patients and outpatients at a single tertiary care medical center (University of California, San Francisco), and **(D)** samples with viruses of B.1.427 or B.1.429 lineage. Note that a Ct difference of 1 represents a 2-fold difference in the virus concentration (Drew et al., 2020). The solid horizontal line in the center box denotes the mean value. $****p < 0.0001$; $***p < 0.001$; $**p < 0.01$; $*p < 0.05$; NS, non-significant. Welch's t-test was used to determine significance.

Figure 4. Increased infectivity of L452R-carrying pseudoviruses.

(A) Upper panel: ribbon diagram of the SARS-CoV-2 spike RBD in cyan bound to ACE2 receptor in magenta (PDB: 6M0J). The receptor-binding motif of RBD is colored in dark cyan with L452 in solid spheres and F490 and L492 with dotted spheres. Sugars and Zn2+ are shown in gray. The position of N501 in direct contact with the ACE2 receptor is also shown for purposes of comparison. **Lower panel**: surface representation of the spike RBD showing the hydrophobic patch outlined by L452, F490, and L492.

(B) Levels of infection of SARS-CoV-2 spike pseudoviruses carrying D614G alone or D614G with N501Y, L452R, or W152C mutations in 293T cells stably expressing ACE2 and TMPRSS2. 293T cells were seeded in 96-well plates and infected with high (6 ng, **left**) or low (3 ng, **right**)

concentrations of the indicated pseudoviruses for 48 h. Two biological replicates were assessed in two independent experiments, with 3 technical replicates per experiment.

(C) Levels of infection in human lung airway organoids (HAO) stably expressing ACE2. HAO were seeded in 24-well plates and infected with high (4 ng, **left**) or low (2 ng, **right**) concentrations of the indicated pseudoviruses for 72 h. Pseudovirus cell entry was measured with a luciferase assay. The error bars represent the SD of 3 technical replicates. Dunn's multiple comparisons test was used to determine significance. Note that each of the N501Y, L452R, and W152G pseudoviruses also carries D614G.

NS, not significant.

Figure 5. B.1.427/B.1.429 variant resistance to antibody neutralization in vitro.

(A) Antibody neutralization titers from 9 convalescent patients and 12 vaccine recipients against cultured WA1 (control), D614G (control), and B.1.429 viral isolates were assessed using a PRNT assay. Lines connect the individual plasma samples tested pairwise for neutralization **(top row)**. Only a subset of the plasma samples were tested with the WA1 and D614G head-to-head comparisons **(top row, right)**. The dotted lines denote the upper and lower bounds for the PRNT assay (1:100 to 1:3,200). Plasma samples that did not exhibit detectable neutralizing activity at titers above the lower threshold are shown as transparent. Individual PRNT50 measurements are plotted along with error bars denoting the median and SD **(bottom row).**

(B) Antibody neutralization titers from 10 convalescent patients against cultured WA1 (control), D614G (control), and B.1.427 viral isolates were assessed by 50% CPE endpoint dilution. Lines connect the individual plasma samples tested pairwise for neutralization **(top row)**. Individual TCID50 measurements are plotted along with error bars denoting the median and SD **(bottom row)**. A Wilcoxon matched pairs signed-rank test was used to determine significance.

NS, not significant; PRNT, plaque-reduction neutralization test; CPE, cytopathic effect; TCID, tissue culture infective dose.Abbreviations: NS, not significant; PRNT, plaque-reduction neutralization test; CPE, cytopathic effect; TCID, tissue culture infective dose.

Figure 1

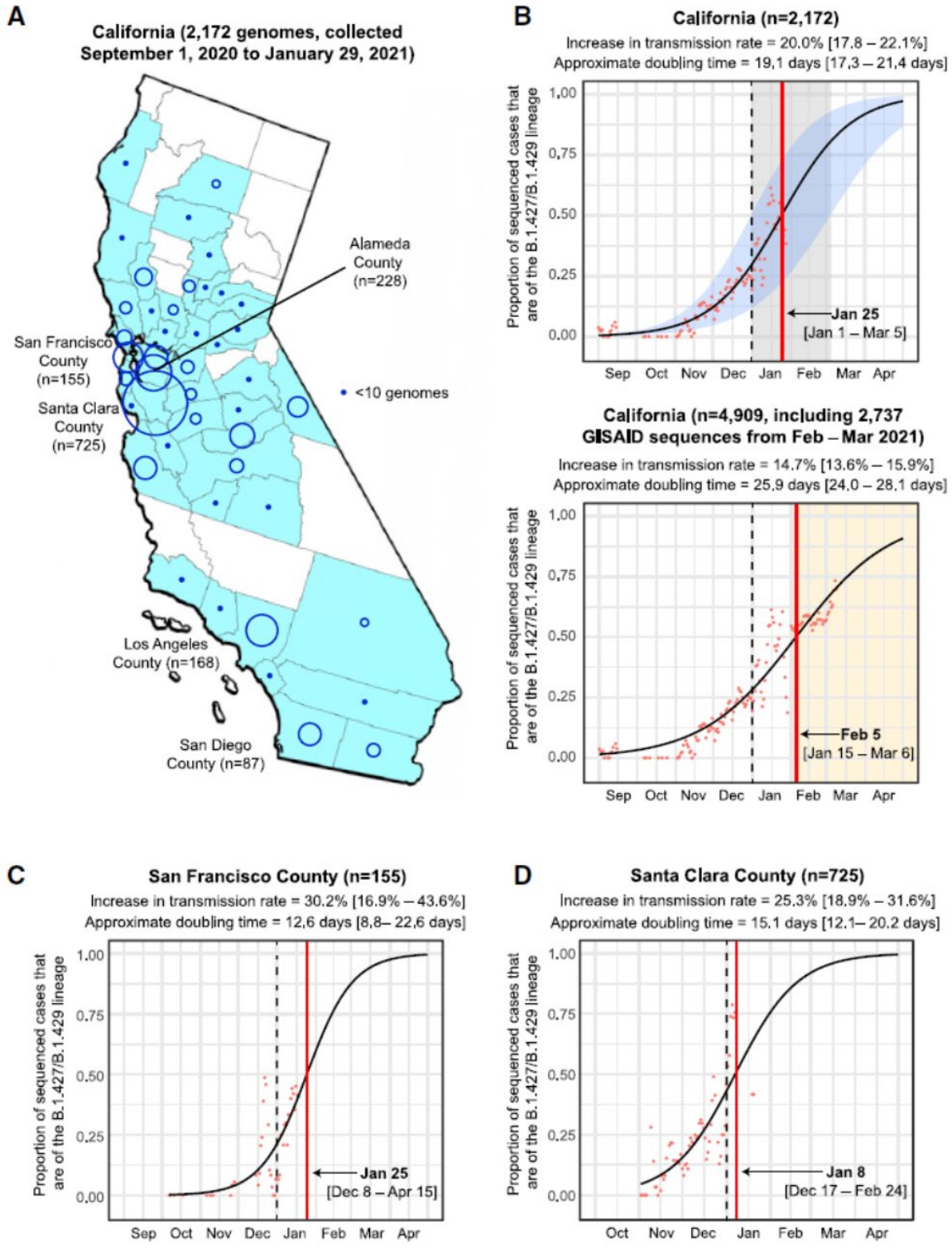

A
California (2,172 genomes, collected September 1, 2020 to January 29, 2021)
Alameda County (n=228)
San Francisco County (n=155)
Santa Clara County (n=725)
<10 genomes
Los Angeles County (n=168)
San Diego County (n=87)
B
California (n=2,172)
Increase in transmission rate = 20.0% [17.8 – 22.1%]
Approximate doubling time = 19,1 days [17,3 – 21,4 days]
Proportion of sequenced cases that are of the B.1.427/B.1.429 lineage
Jan 25
[Jan 1 – Mar 5]
California (n=4,909, including 2,737 GISAID sequences from Feb – Mar 2021)
Increase in transmission rate = 14.7% [13.6% – 15.9%]
Approximate doubling time = 25.9 days [24.0 – 28.1 days]
Feb 5
[Jan 15 – Mar 6]
C
San Francisco County (n=155)
Increase in transmission rate = 30.2% [16.9% – 43.6%]
Approximate doubling time = 12,6 days [8,8– 22,6 days]
Jan 25
[Dec 8 – Apr 15]
D
Santa Clara County (n=725)
Increase in transmission rate = 25.3% [18.9% – 31.6%]
Approximate doubling time = 15.1 days [12.1– 20.2 days]
Jan 8
[Dec 17 – Feb 24]
Sep Oct Nov Dec Jan Feb Mar Apr

Figure 2

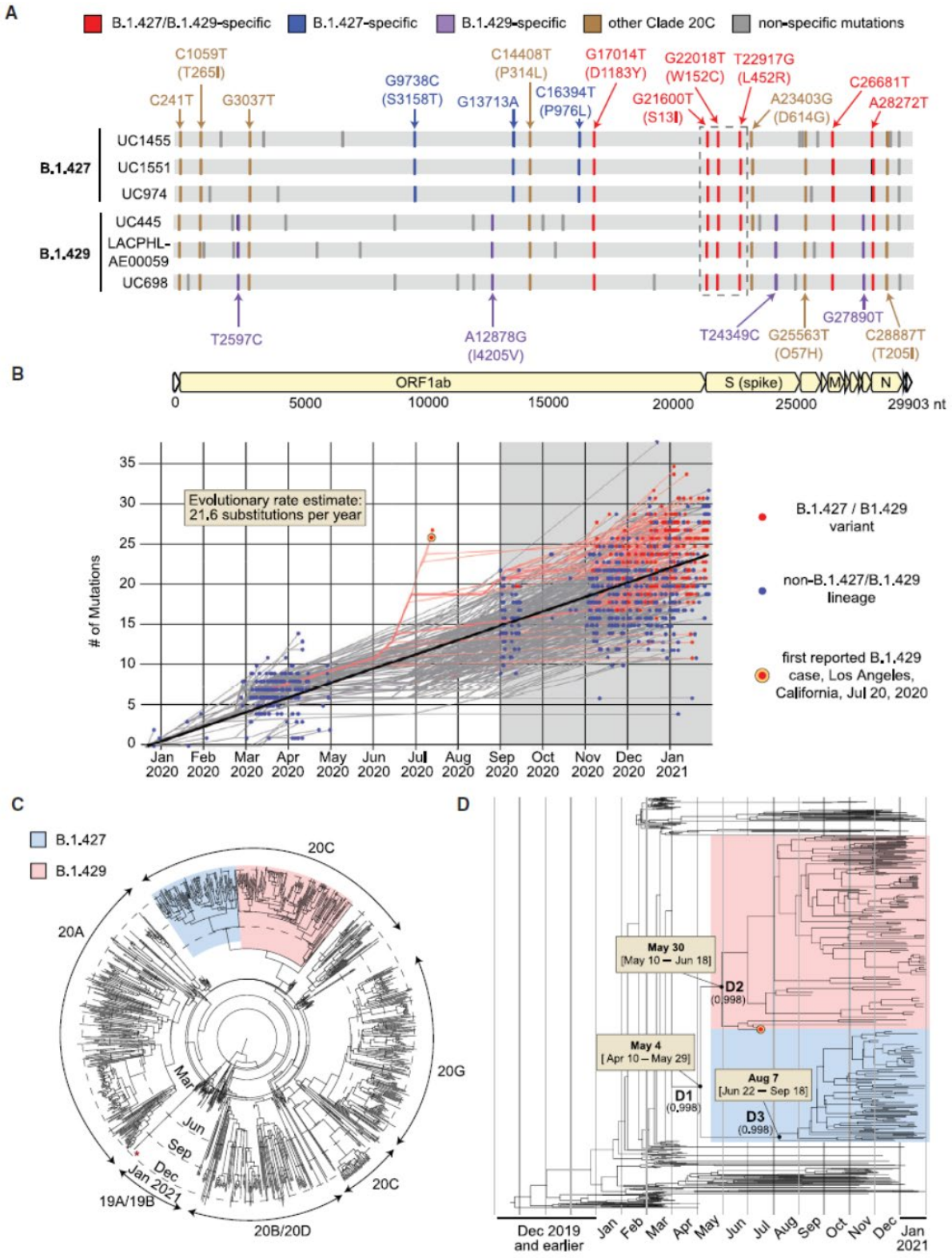

A
B.1.427/B.1.429-specific
B.1.427-specific
B.1.429-specific
other Clade 20C
non-specific mutations
C1059T (T265I)
C241T
G3037T
G9738C (S3158T)
C14408T (P314L)
G13713A
C16394T (P976L)
G17014T (D1183Y)
G22018T (W152C)
T22917G (L452R)
G21600T (S13I)
A23403G (D614G)
C26681T
A28272T
B.1.427
UC1455
UC1551
UC974
B.1.429
UC445
LACPHL-AE00059
UC698
T2597C
A12878G (I4205V)
T24349C
G25563T (O57H)
G27890T
C28887T (T205I)
B
ORF1ab
S (spike)
M
N
0
5000
10000
15000
20000
25000
29903 nt
Evolutionary rate estimate: 21.6 substitutions per year
# of Mutations
Jan 2020
Feb 2020
Mar 2020
Apr 2020
May 2020
Jun 2020
Jul 2020
Aug 2020
Sep 2020
Oct 2020
Nov 2020
Dec 2020
Jan 2021
B.1.427 / B1.429 variant
non-B.1.427/B.1.429 lineage
first reported B.1.429 case, Los Angeles, California, Jul 20, 2020
C
B.1.427
B.1.429
20C
20A
20G
20C
20B/20D
19A/19B
Mar
Jun
Sep
Dec
Jan 2021
D
May 30 [May 10 – Jun 18]
D2 (0.998)
May 4 [Apr 10 – May 29]
D1 (0.998)
Aug 7 [Jun 22 – Sep 18]
D3 (0.998)
Dec 2019 and earlier
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan 2021

Figure 3

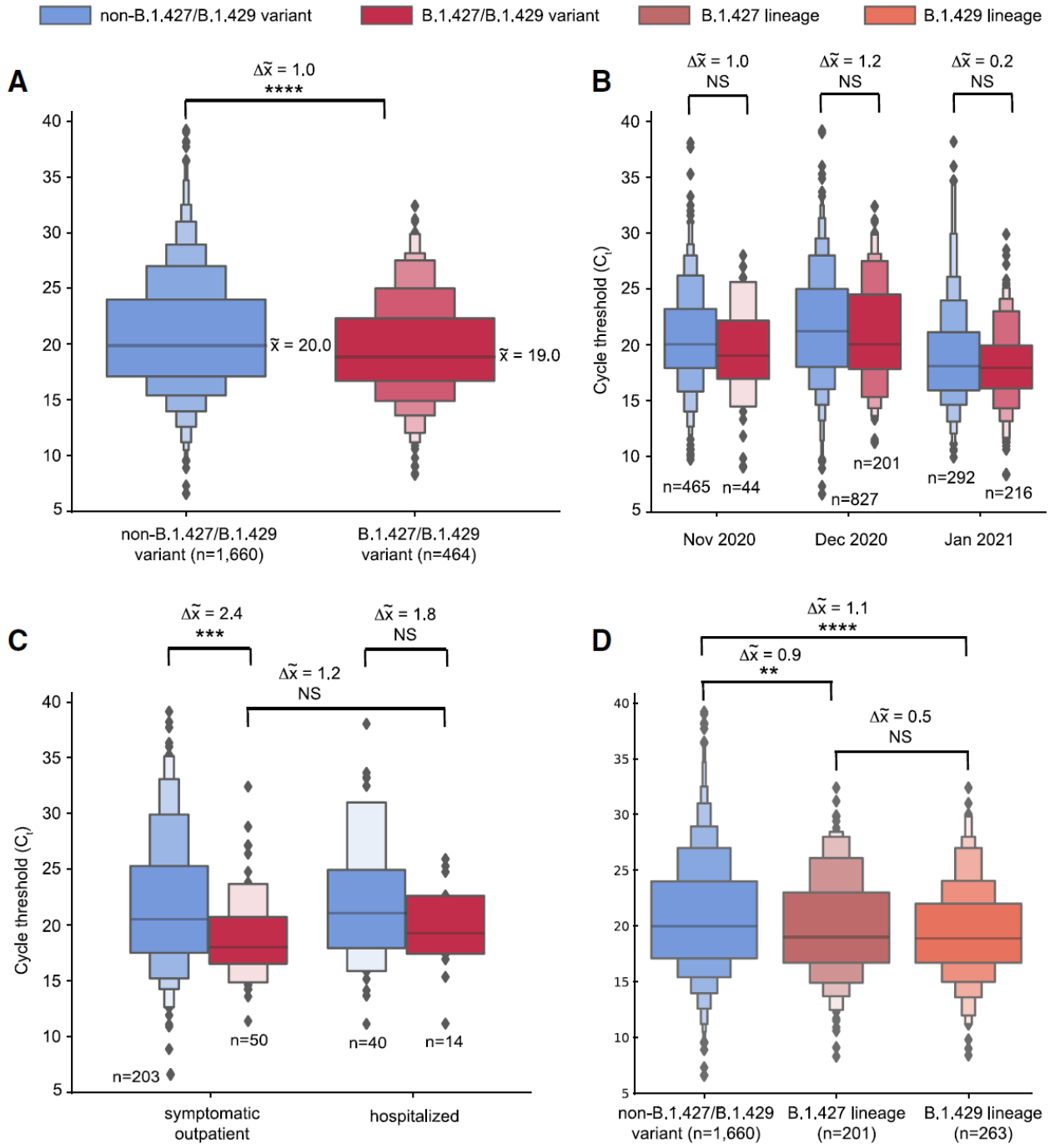

non-B.1.427/B.1.429 variant
B.1.427/B.1.429 variant
B.1.427 lineage
B.1.429 lineage
A
Δx̃ = 1.0
****
x̃ = 20.0
x̃ = 19.0
non-B.1.427/B.1.429 variant (n=1,660)
B.1.427/B.1.429 variant (n=464)
B
Δx̃ = 1.0
NS
Δx̃ = 1.2
NS
Δx̃ = 0.2
NS
Cycle threshold (Ct)
n=465
n=44
n=827
n=201
n=292
n=216
Nov 2020
Dec 2020
Jan 2021
C
Δx̃ = 2.4
***
Δx̃ = 1.8
NS
Δx̃ = 1.2
NS
Cycle threshold (Ct)
n=203
n=50
n=40
n=14
symptomatic outpatient
hospitalized
D
Δx̃ = 1.1
****
Δx̃ = 0.9
**
Δx̃ = 0.5
NS
non-B.1.427/B.1.429 variant (n=1,660)
B.1.427 lineage (n=201)
B.1.429 lineage (n=263)

Figure 4

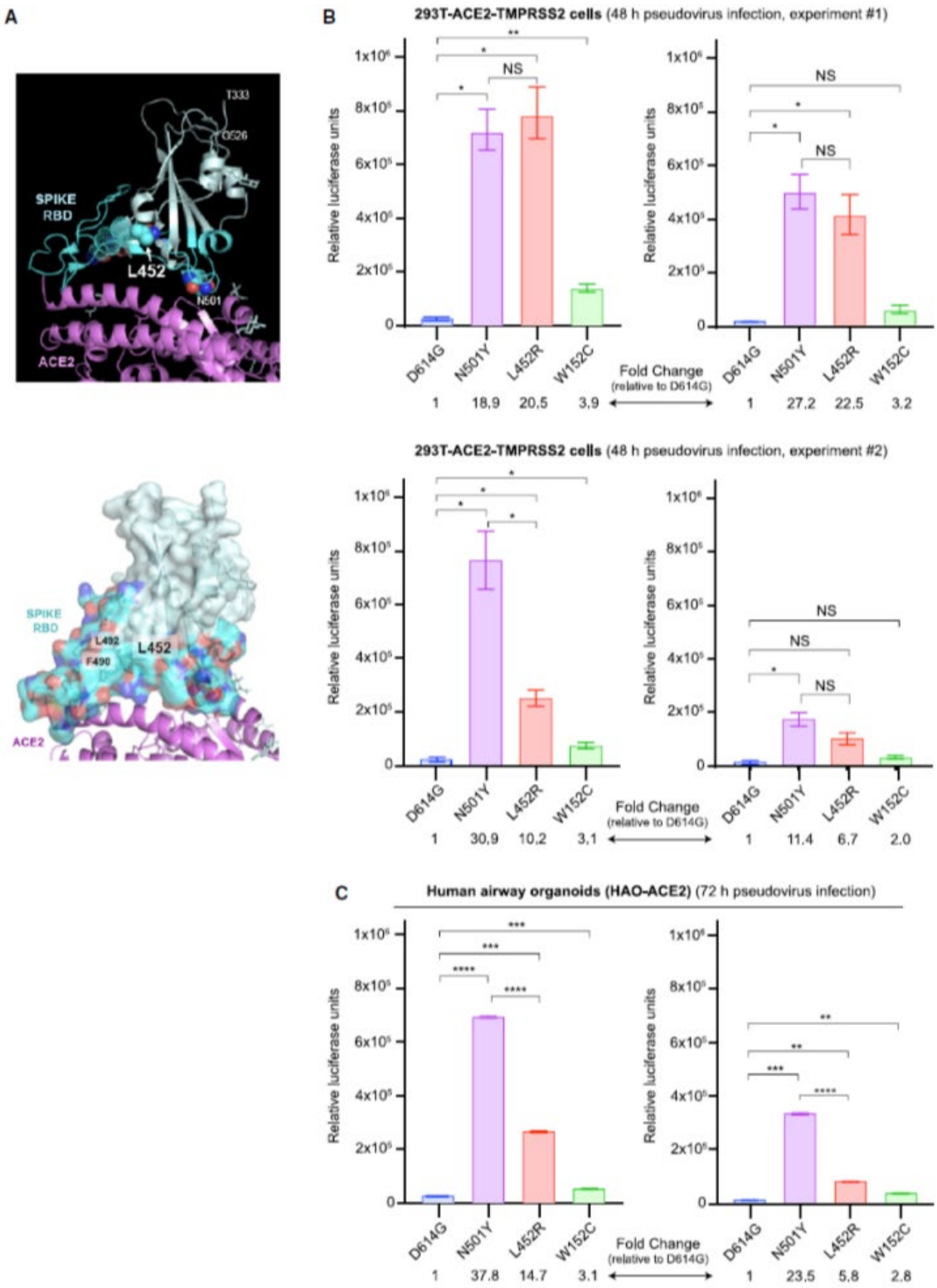
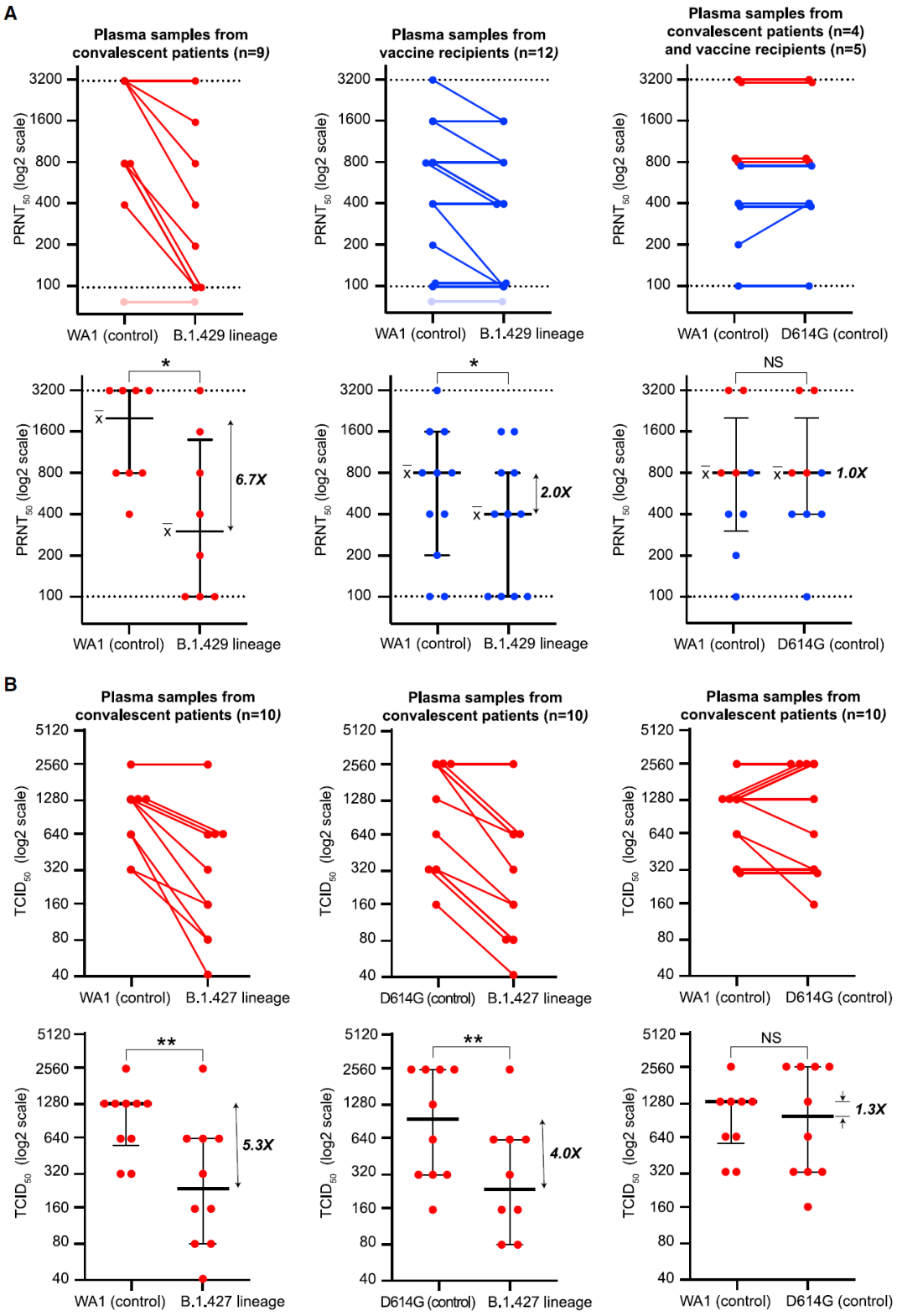

A
T333
O526
SPIKE
RBD
L452
N501
ACE2
SPIKE
RBD
L492
F490
L452
ACE2
B
293T-ACE2-TMPRSS2 cells (48 h pseudovirus infection, experiment #1)
Relative luciferase units
1x10⁶
8x10⁵
6x10⁵
4x10⁵
2x10⁵
0
NS
D614G
N501Y
L452R
W152C
Fold Change
(relative to D614G)
1
18,9
20,5
3,9
1
27.2
22.5
3.2
293T-ACE2-TMPRSS2 cells (48 h pseudovirus infection, experiment #2)
1
30,9
10,2
3,1
1
11.4
6.7
2.0
C
Human airway organoids (HAO-ACE2) (72 h pseudovirus infection)
1
37.8
14.7
3.1
1
23,5
5,8
2.8

Figure 5


A
Plasma samples from convalescent patients (n=9)
Plasma samples from vaccine recipients (n=12)
Plasma samples from convalescent patients (n=4) and vaccine recipients (n=5)
PRNT50 (log2 scale)
3200
1600
800
400
200
100
WA1 (control)
B.1.429 lineage
D614G (control)
*
NS
6.7X
2.0X
1.0X
B
Plasma samples from convalescent patients (n=10)
TCID50 (log2 scale)
5120
2560
1280
640
320
160
80
40
B.1.427 lineage
**
5.3X
4.0X
1.3X

Acknowledgments

We acknowledge the help from Delsy Martinez and Tyler Miyasaki at UCSF CAT core facility for genome sequencing efforts using NovaSeq. We acknowledge Maria Salas, Elizabeth Baylis and the entire COVIDNet team at the Viral and Rickettsial Disease Laboratory of the California Department of Public Health for their support of our community viral WGS efforts. We thank the Whelan lab at the Washington University School of Medicine for the Vero TMPRSS2 cell line. We thank Satish Pillai and Hannah S. Sperbert at Vitalant Research Institute (San Francisco, CA) for providing the 293T-ACE2-TMPRSS2 stable cell line used for pseudovirus entry assays. We thank Mehrdad Arjomandi at the San Francisco Veterans Administration Medical Center for providing the clinical samples to generate the human alveolar lung organoids. We thank Dr. Kristian Andersen and his laboratory at the Scripps Research Institute for generating code (available on Github) to calculate logistic growth models of SARS-CoV-2 infection based on positive tests over time.

We gratefully acknowledge the authors from the originating laboratories and the submitting laboratories who generated and shared via GISAID genetic sequence data from an additional 2,737 genomes (from samples from California collected February 1 to March 11, 2021) on which this research is based (**Table S5).**

The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views or opinions of the California Department of Public Health or the California Health and Human Services Agency.

Funding

This work has been funded by a Laboratory for Genomics Research (LGR) Excellence in Research Award (LFL), a Fast Grant from Emergent Ventures (SKW), the Innovative Genomics Institute (CYC, MO, LFL, SW, PF, HS), the New Frontiers in Research Fund provided by the Canadian Institutes of Health Research (CYC), the Roddenberry Foundation (MO), and NIH grants R33-AI129455 (CYC) and 5DP1DA038043 (MO).

Author contributions

CYC, MO, and RA conceived and designed the study. CYC, XD, MAG-K, VS, CW, and GRK coordinated the sequencing efforts and laboratory studies. XD, MAG-K, MMK, VS, CW, AS-G, DRG, KRR, CSSM, BS, P-YC, US-G, TYT, JMH, CRS, PVL, YX, and MKM performed experiments. CYC, SF, and XD assembled and curated viral genomes. CYC performed the

phylogenetic and molecular clock analyses. CYC, XD, MAG-K, VS, CW, KRR, ASG, NPR, JB, JT, JC, GRK, and CYC analyzed data. VS, CW, AS-G, ASG, NPR, KRR, JAS, and SM collected and sequenced SARS-CoV-2 samples from UCSF and throughout California. PH and NMG collected and sequenced samples from Los Angeles County. CA and DF collected and sequenced samples from Monterey County. FL, PAF, HS, and SKW collected and sequenced samples from Alameda County. CYC, XD, MAGK, VS, and CW wrote the manuscript. CYC, MAGK, GRK, and VS prepared the figures. CYC revised the figures. CYC, XD, MAGK, VS, DAW, JKH, and CW edited the manuscript. All authors read the manuscript and agree to its contents.

Declaration of interests

CYC. receives support for SARS-CoV-2 research unrelated to this study from Abbott Laboratories and Mammoth Biosciences. The other authors declare no competing interests.

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Charles Chiu (charles.chiu@ucsf.edu).

Materials availability

Passaged aliquots of cultured SARS-CoV-2 B.1.427 and B.1.429 viruses, pseudoviruses bearing the D614G, L452R, and/or W152C viruses, and SARS-CoV-2 nasal swab / nasopharyngeal samples and/or RNA extracts are available upon request.

Data and code availability

Assembled SARS-CoV-2 genomes in this study were uploaded to GISAID (Elbe and Buckland-Merrett, 2017; Shu and McCauley, 2017) (accession numbers in Table S1) and can be visualized in Nextstrain. Viral genomes were also submitted to the National Center for Biotechnology Information (NCBI) GenBank database (accession numbers MW972466-MW974550), BioProject accession number PRJNA722044 and umbrella BioProject accession number PRJNA171119, Chiu laboratory at UCSF; umbrella BioProject accession number PRJNA639591, Wyman laboratory at UC Berkeley). FASTA files, XML files, and scripting code used for SARS-CoV-2 genome assembly and phylogenetic / molecular dating analyses are available in a Zenodo data repository (https://doi.org/10.5281/zenodo.4688394) (Chiu and Servellita, 2021).

Experimental model and subject details

Human sample collection and ethics

Remnant nasal/nasopharyngeal (N/NP) swab samples in universal transport media (UTM) or viral transport media (VTM) (Copan Diagnostics, Murrieta, CA, USA) from RT-PCR positive COVID-19 patients were obtained from the University of California, San Francisco (UCSF) Clinical Microbiology Laboratory, the Innovative Genomics Institute (IGI) at University of California, Berkeley, California Department of Public Health, Santa Clara County and Los Angeles County for SARS-CoV-2 genome sequencing. A small fraction of swab samples (< 1%) were obtained from the anterior nares. Clinical samples from state and county public health laboratories were collected and sequenced as part of routine public health surveillance activities. Clinical samples from the IGI were sequenced under a waiver from the UC Berkeley Office for the Protection of Human Subjects. Clinical samples from UCSF were collected for a biorepository and sequenced

according to protocols approved by the UCSF Institutional Review Board (protocol number 10-01116, 11-05519). Samples were obtained from pediatric and adult donors of all genders. No analyses based on sex or age were conducted in this study.

Cell culture models

Cells used for this study include Vero E6 cells, Vero-81 cells, Vero cells overexpressing human TMPRSS2 (Vero-TMPRSS2), A549 cells stably expressing ACE2 (A549-ACE2), and 293T cells stably expressing ACE2 and TMPRSS2 (293T-ACE2-TMPRSS2) (Khanna et al., 2020).

. Vero E6 cells were cultured in MEM supplemented with 1x penicillin-streptomycin-glutamine (GIBCO) and 10% fetal calf serum (FCS). Vero-81 cells were cultured with MEM supplemented with 1x penicillin-streptomycin (GIBCO) and glutamine (GIBCO) and 5% FCS (Hyclone). Vero-TMPRSS2 cells were maintained in DMEM supplemented with 1x sodium pyruvate, 1x penicillin-streptomycin-glutamine and 10% FCS. A549-ACE2 cells were cultured in DMEM/F-12 media supplemented with 10% FCS. 293T-ACE2-TMPRSS2 cells were cultured in DMEM supplemented with 10% FCS, 1% penicillin-streptomycin, 10 µg/mL blasticidin and 1 µg/mL puromycin. Cell cultures were maintained in a humidified incubator at 37°C in 5% CO2 in the indicated media and passaged every 3-4 days.

Human airway lung organoids (HAO)

Human airway lung organoids (HAO) were grown from whole-lung lavages from adult donors and cultured as previously reported (Sachs et al., 2019). Briefly, single cells were suspended in 65% reduced growth factor BME2 (Basement Membrane Extract, Type 2). From this mixture, 50 µL drops with 1,000–40,000 cells were seeded in 24-well suspension culture plates to generate three-dimensional organoids representing the 4 major epithelial cell types (basal cells, club cells, goblet cells, and ciliated cells). In order to generate HAO stably expressing ACE2 (HAO-ACE2), organoids were transduced with lentiviruses encoding ACE2 for 6 hours, expanded for 48 hours, and selected with blasticidin (1 µg/ml) for 7 days.

Isolation of SARS-CoV-2 viral strains for neutralization studies

For the B.1.429 neutralization studies, a non-B.1.427/B.1.429 variant SARS-CoV-2/human/USA/CA-UCSF-0001C/2020 clinical isolate carrying the D614G spike mutation was cultured as previously described (Samuel et al., 2020) and passaged in A549-ACE2 expressing cells. For isolation of the B.1.429 lineage virus, 100 µL of a NP swab sample from a COVID-19 patient that was previously sequenced and identified as B.1.429 was mixed 1:1 with serum free

DMEM (supplemented with 1x sodium pyruvate and 1x penicillin-streptomycin-glutamine), and two-fold serial dilutions were made of the sample over six wells of a 96-well plate. 100 µL of freshly trypsinized Vero-TMPRSS2 cells resuspended in DMEM (supplemented with 1x sodium pyruvate, 2x penicillin-streptomycin-glutamine, 5 µg/mL amphotericin B and 10% FCS) was added to each well and mixed. The culture was incubated at 37°C in 5% CO2 for 4-6 days and cytopathic effect (CPE) on cells was evaluated daily using a light microscope. The contents of wells positive for CPE were collected and stored at −80°C as a passage 0 stock (P0). P1 stocks were made following infection of four near confluent wells of a 24-well plates with Vero-TMPRSS2 using the P0 stock. Supernatants were harvested 48 hours later after centrifugation at 800 g for 7 minutes. P2 stocks were similarly made after infection of a near confluent T25 plate seeded with Vero E6 cells. All steps for isolation of the B.1.429 lineage virus were done in a Biosafety Level 3 lab using protocols approved by the Institutional Biosafety Committee at UCSF.

For the B.1.427 neutralization studies, B.1.427 and non-B.1.427/B.1.429 variant D614G viruses were cultured from NP swab samples from COVID-19 patients identified by viral whole-genome sequencing as being infected by the B.1.427 or non-B.1.427/B.1.429 variant D614G lineage. Briefly, 100 µL of NP swab sample was diluted 1:5 in PBS supplemented with 0.75% bovine serum albumin (BSA-PBS) and added to confluent Vero-81 cells in a T25 flask. After adsorption for 1 h, additional media was then added, and the flask was incubated at 37°C with 5% CO2 for 3-4 days with daily monitoring for CPE. The contents were collected, clarified by centrifugation and stored at −80C as passage 0 stock. P1 stock was made by inoculation of Vero-81 confluent T150 flasks with 1:10 diluted p0 stock and similarly monitored and harvested to approximately 50% confluency. All steps for isolation of the B.1.427 lineage virus were done in a Biosafety Level 3 lab at the Viral and Rickettisial Disease Laboratory (VRDL) at the California Department of Public Health (CDPH).

For both the B.1.429 and B.1.427 neutralization studies, the SARS-CoV-2 USA-WA1/2020 strain (BEI resources) was passaged in Vero E6 cells or Vero-81 cells and used as a control. All stocks were resequenced and the consensus assembled viral genomes were identical to the genomes derived from the primary NP samples and carried all of the expected mutations.

Method details

SARS-CoV-2 diagnostic testing

Due to variation in results reported by different clinical testing platforms used at UCSF, the Taqpath Multiplex Real-time RT-PCR test, which includes nucleoprotein (N) gene, spike (S) gene, and orf1ab gene targets, was used to determine cycle threshold (Ct) values for PCR-positive samples. The Taqpath assay was also used for determining Ct values for PCR-positive samples from Alameda County that were sequenced by the University of California, Berkeley IGI and from the California Department of Public Health.

SARS-CoV-2 genome sequencing

NP swab samples were prepared using 100 uL of the primary sample in UTM or VTM mixed with 100uL DNA/RNA shield (Zymo Research, #R1100-250). The 1:1 sample mixture was then extracted using the Omega BioTek MagBind Viral DNA/RNA Kit (Omega Biotek, #M6246-03) on KingFisherTM Flex Purification System with a 96 deep-well head (ThermoFisher, 5400630). Extracted RNA was reverse transcribed to complementary DNA and tiling multiplexed amplicon PCR was performed using SARS-CoV-2 primers Version 3 according to a published protocol (Quick et al., 2017). Amplicons were ligated with adapters and incorporated with barcodes using NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, #E7645L). Libraries were barcoded using NEBNext Multiplex Oligos for Illumina (96 unique dual-index primer pairs) (New England Biolabs, #E6440L) and purified with AMPure XP beads (Beckman-Coulter, #63880). Amplicon libraries were then sequenced on either Illumina MiSeq or Novaseq 6000 as 2x150 paired-end reads (300 cycles).

Viral genome assembly and variant calling

Genome assembly of viral reads and variant calling were performed using an in-house developed bioinformatics pipeline as previously described (Deng et al., 2020). In short, Illumina raw paired-end reads were first screened for SARS-CoV-2 sequences using BLASTn (BLAST+ package 2.9.0) alignment against viral reference genome NC_045512, and then processed using the BBTools suite, v38.87 (Bushnell, 2021). Adaptor sequences were trimmed and low-quality reads were removed using BBDuk, and then mapped to the NC_045512 reference genome using BBMap. Variants were called with CallVariants and a depth cutoff of 5 was used to generate the final assembly. A genome coverage breadth of ≥ 70% was required for inclusion in the study.

PANGOLIN (Phylogenetic Assignment of Named Global Outbreak LINeages) v.2.3.8 was used to assign SARS-CoV-2 lineages (Rambaut et al., 2020a).

Multiple sequence alignment of 6 B.1.427/B.1.429 genomes and the Wuhan Hu-1 prototypical genome (GISAID ID: EPI_ISL_402125, GenBank: MN908947) was performed using the MAFFT aligner v7.388 (Katoh and Standley, 2013) as implemented in Geneious v11.1.5 (Kearse et al., 2012).

Phylogenetic analysis

High-quality SARS-CoV-2 genomes (n = 2,519, 2,172 generated in the current study and 347 used as representative global genomes) were downloaded from the Global Initiative on Sharing of All Influenza Data (GISAID) database and processed using the Nextstrain bioinformatics pipeline Augur using IQTREE v1.6. Branch locations were estimated using a maximum-likelihood discrete traits model. The resulting tree consisting of 1,153 subsampled genomes was visualized in the Nextstrain web application Auspice (root-to-tip divergence plot in Figure 2B) and in Geneious v11.1.5 (circular phylogenetic tree in Figure 2C) (Kearse et al., 2012).

Molecular dating analysis of SARS-CoV-2 for estimating the TMRCA (time to most recent common ancestor) and divergence dates for the B.1.426/B.1.427 variant was performed using the Markov chain Monte Carlo (MCMC) method implemented by Bayesian Evolutionary Analysis on Sampling Trees (BEAST) software v.2.63 (Bouckaert et al., 2019; Drummond et al., 2012). To decrease computational turnaround time, a representative subset of 490 out of the 1,153 subsampled genomes was identified by combining 225 of the 227 B.1.427/B.1.429 genomes, 100 randomly selected non-B.1.427/B.1.429 variant genomes from California, and all 165 global sequences outside of California. Two B.1.427/B.1.429 genomes (UC1504 and UC464) were found to be outliers that did not map to the B.1.427/B.1.429 phylogenetic cluster due to regions of low genomic coverage and were removed from further analysis. BEAST analysis of the 490 representative genomes was performed using an HKY85 nucleotide substitution model with a strict clock and exponential population growth (Laplace distribution). All models were run using default priors. The chain length was set to 100 million states with a 10% burn-in. Convergence was evaluated using Tracer v1.7.1 (Rambaut et al., 2018). As a single BEAST run resulted in some parameters with effective sample size (ESS) values of < 200, the logged MCMC output of two runs, each consisting of 100 million states, was combined using LogCombiner v.1.10.4 from the BEAST package. The two runs were inspected prior to combining them and were found to yield nearly identical tree topologies. After combining the MCMC chains from both runs, the ESS

values for all parameters were > 200, ranging from 265 to 13,484. The resulting maximum clade credibility (MCC) tree was generated using TreeAnnotator v.2.6.3 (Drummond et al., 2012) and visualized using FigTree v.1.4.4 (Rambaut, 2021).

SARS-CoV-2 receptor binding domain mutagenesis and pseudovirus infection assay

SARS-CoV-2 spike mutants (D614G, D614G+W152C, D614G+L452R, and D614G+N501Y) were cloned using standard site-directed mutagenesis and PCR. Pseudoviruses typed with these spike mutants were generated as previously described with modifications (Crawford et al., 2020). Briefly, 293T cells were transfected with plasmid DNA (per 6-well plate: 340 ng of spike mutants, 1 µg CMV-Gag-Pol (pCMV-dΔR8.91), 125 ng pAdvantage (Promega), 1 µg Luciferase reporter) for 48 h. Supernatant containing pseudovirus particles was collected, filtered (0.45 µm), and stored in aliquots at −80°C. Pseudoviruses were quantified with a p24 assay (Takara #632200), and normalized based on titer prior to infection for entry assays.

Human airway organoids (HAO) stably expressing ACE2 (HAO-ACE2) or 293T cells stably expressing ACE2 and TMPRSS2 (293T-ACE2-TMPRSS2) were infected with an equivalent amount of the indicated pseudoviruses in the presence of 5-10 ug/ml of polybrene for 72h. Pseudovirus entry was assayed using a luciferase assay (Promega #E1501) and luminescence was measured in a plate reader (TECAN, Infinite 200 Pro M Plex). Two independent experiments were run for the 293T pseudovirus assays (2 biological replicates), with 3 technical replicates run per experiment. The HAO pseudovirus assays were run as a single experiment with 3 technical replicates.

Plaque reduction neutralization tests using a B.1.429 lineage virus

Conventional PRNT assays were done using P2 stocks of B.1.429 lineage viruses and the USA-WA1/2020 isolate passaged on Vero E6 cells. Patient plasma was heat inactivated at 56°C for 30 minutes, clarified by centrifugation at 10,000 relative centrifugal force (rcf.) for 5 minutes and aliquoted to minimize freeze thaw cycles. Serial 2-fold dilutions were made of plasma in PBS supplemented with 0.75% bovine serum albumin (BSA). Plasma dilutions were mixed with ~100 plaque forming units (pfu) of viral isolates in serum free MEM in a 1:1 ratio and incubated for 1 hr at 37°C. Final plasma dilutions in plasma-virus mixtures ranged from 1:100 to 1:3200. 250 µL of plasma-virus mixtures were inoculated on a confluent monolayer of Vero E6 cells in 6-well plates, rocked and incubated for 1 h in a humidified incubator at 37°C in 5% CO2. After incubation, 3 mL of a mixture of MEM containing a final concentration of 2% FCS, 1x penicillin-streptomycin-glutamine and 1% melted agarose, maintained at ~56°C, was added to the wells. After 72 h of

culture as above, the wells were fixed with 4% paraformaldehyde for 2 h, agarose plugs were removed, and wells were stained with 0.1% crystal violet solution. Plaques were counted and the PRNT50 values were defined as the serum dilution at which 50% or more of plaques were neutralized. Assays were done in duplicate, and a positive control and negative control were included using plasma with known neutralizing activity (diluted 1:50) and from a SARS-CoV-2 unexposed individual (1:20 dilution), respectively. All steps were done in a Biosafety Level 3 lab using protocols approved by the Institutional Biosafety Committee at UCSF.

CPE endpoint neutralization assays using a B.1.427 lineage virus

CPE endpoint neutralization assays were done following the limiting dilution model (Wang et al., 2005) and using P1 stocks of B.1.427, D614G, and USA-WA1/2020 lineages. Convalescent patient plasma was diluted 1:10 and heat inactivated at 56°C for 30 min. Serial 2-fold dilutions of plasma were made in BSA-PBS. Plasma dilutions were mixed with 100 TCID50 of each virus diluted in BSA-PBS at a 1:1 ratio (220 µL plasma dilution and 220 µL virus input) and incubated for 1 hour at 37C. Final plasma dilutions in plasma-virus mixture ranged from 1:20 to 1:2560. 100 µL of the plasma-virus mixtures were inoculated on confluent monolayer of Vero-81 cells in 96-well plates in quadruplicate and incubated at 37°C with 5% CO2 incubator. After incubation 150 µL of MEM containing 5% FCS was added to the wells and plates were incubated at 37°C with 5% CO2 until consistent CPE was seen in virus control (no neutralizing plasma added) wells. Positive and negative controls were included as well as cell control wells and a viral back titration to verify TCID50 viral input. Individual wells were scored for CPE as having a binary outcome of ‘infection” or ‘no infection’. The TCID50 was calculated as the dose that produced cytopathic effect in > 50% of the inoculated wells. All steps were done in a Biosafety Level 3 lab using approved protocols.

Data visualization

The plots in Figure S1 were generated using graphical visualization tools at outbreak.info (Gangavarapu et al., 2020) and Microsoft Excel v16.47.1 and edited in Adobe Illustrator 23.1.1, using data from GISAID (Elbe and Buckland-Merrett, 2017; Shu and McCauley, 2017) and the California COVID-19 data tracker (CDPH (California Department of Public Health), 2021b). Other figures were generated using R v4.0.3 and Python v3.7.10 and edited in Adobe Illustrator.

Quantification and statistical analysis

The proportion of B.1.427/B.1.429 was estimated by dividing the number of B.1.427/B.1.429 variant cases by the total number of samples sequenced at a given location and collection date. A logistic growth curve fitting to the data points was generated using a non-linear least-squares approach, as implemented by the nls() function in R(version 4.0.3), and using code generated by the laboratory of Dr. Kristian Andersen at the Scripps Institute (https://github.com/andersen-lab/paper_2021_early-b117-usa). We estimated the increase in relative transmission rate of the B.1.427/B.1.429 variant by multiplying the logistic growth rate, defined as the change in the proportion of B.1.427/B.1.429 cases per day, by the serial interval, as previously described (Volz et al., 2020; Washington et al., 2021). The serial interval or generation time was defined as the average time taken for secondary cases to be infected by a primary case. The serial interval was found to be linearly proportional to the calculated transmission rate and did not affect the doubling time (Table S4). For SARS-CoV-2 infection, the serial interval has been estimated at 5 – 5.5 days (Rai et al., 2021), 5.5 days (Davies et al., 2021a; Washington et al., 2021), and 6.5 days (Volz et al., 2020); for the data in Figure 1, we used 5.5 days for the serial interval. Similar to the analyses in Washington et al., the doubling time was approximated using the formula: log (2) / logistic growth rate. Outliers corresponding to rolling average date ranges during which only a single B.1.427/B.1.429 variant genome was sequenced (100% proportion of the variant) were removed prior to curve fitting.

Welch's t test, as implemented in R (version 4.0.3) using the rstatix_0.7.0 package and Python (version 3.7.9) using scipy package (version 1.5.2), was used to compare the N gene Ct values between B.1.427/B.1.429 variant and non-B.1.427/B.1.429 groups. For the in vitro pseudovirus infectivity studies, a one-way ANOVA test was used to determine significance. For the PRNT studies, a Wilcoxon matched pairs signed rank test was used to determine significance.

Supplemental Figures and Tables

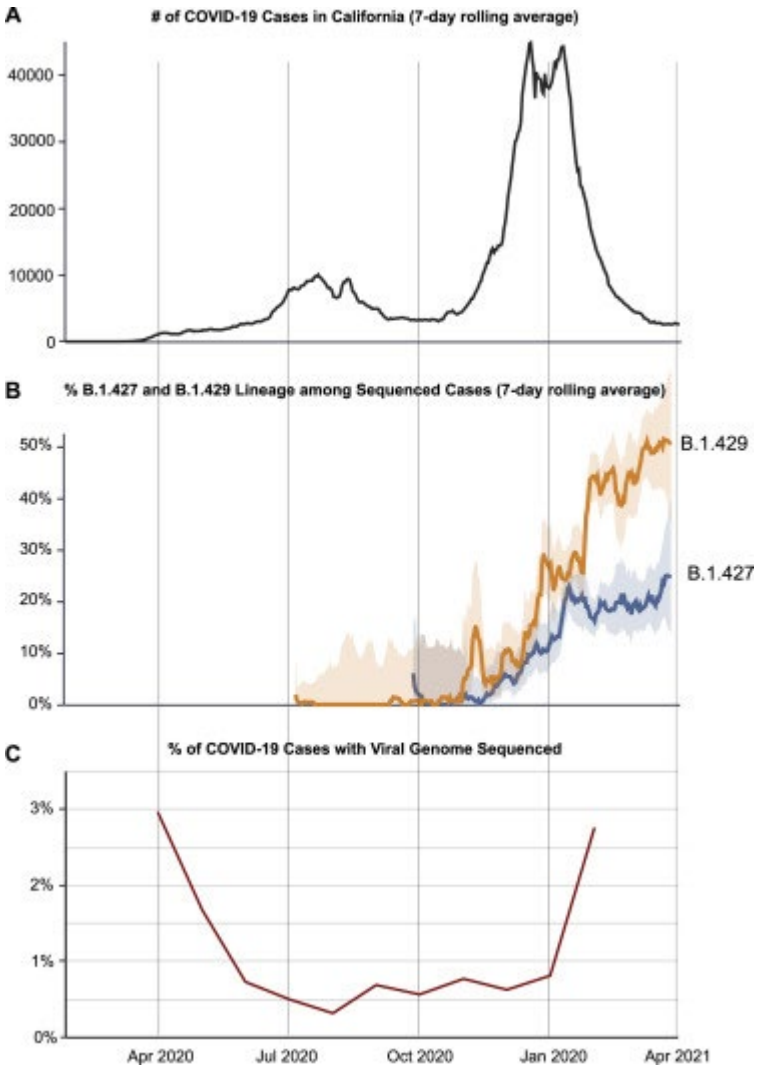


Figure S1. COVID-19 cases, frequency of the B.1.427/B.1.429 variant, and percentage of sequenced cases in California from April 1, 2020 to April 1, 2021, related to Figure 1

(A) Plot showing the reported COVID-19 cases in California. (B) Plot showing the 784 frequency of sequenced cases corresponding to the B.1.427 or B.1.429 variant. (C) Plot 785 showing the % of COVID-19 cases for which the viral genome is sequenced.

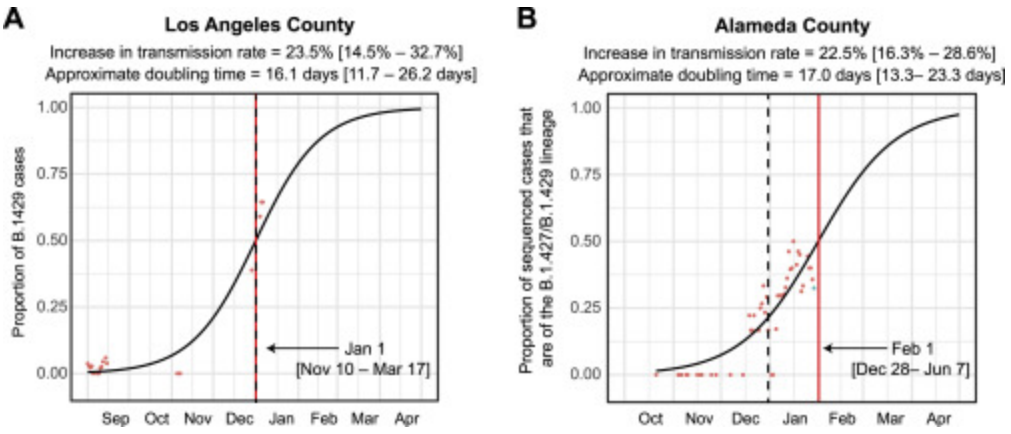


Figure S2. Increasing frequency of the B.1.427/B.1.429 variant in Los Angeles County and Alameda County from September 2020 to January 2021, related to Figure 1

Logistic growth curves fitting the 5-day rolling average of the estimated proportion of B.1.427/B.1.429 variant cases in Los Angeles County (A) and Alameda County (B). A vertical black dotted line is used to denote the transition from 2020 to 2021.

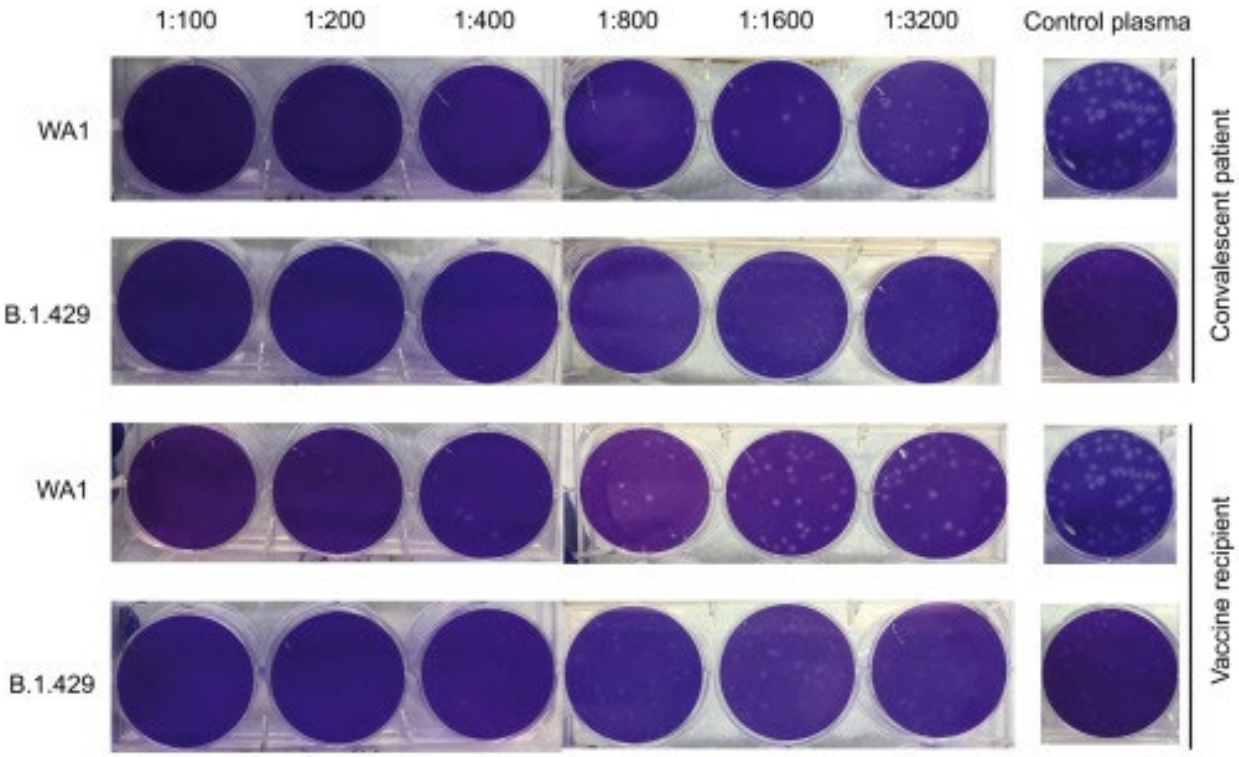


Figure S3. Differential neutralization of WA1 and B.1.429 viruses as measured by plaque-reduction neutralization tests, related to Figure 5

Representative 6-well plates arranged in one line showing viral plaques formed after co-culture with plasma samples from a convalescent patient and vaccine recipient. The same negative control well image is shown in line with the respective viral strain for both vaccine and convalescent samples. The plaques from B.1.429 lineage virus are observed to be small and lighter than those from control WA1 virus. The larger plaques for WA1 are likely due to adaptation in Vero E6 cells; these adaptation mutations have been reported not to impact neutralization responses (Klimstra et al., 2020)..

Table S1. Metadata for the 2,172 genomes analyzed in this study.

"SupplementalTable1.xlsx"

Table S2. List of California counties and the B.1.427/B.1.429 genomes sequenced from each county

Name of county	# of genomes (total n=2,172)	# of B.1.427/B.1.429 genomes (total n=466)	% of B.1.427/B.1.429 genomes
Santa Clara County	725	134	18.5%
Alameda County	228	71	31.1%
Los Angeles County	168	54	32.1%
San Francisco County	155	32	20.6%
Madera County	100	1	1.0%
Contra Costa County	97	31	32.0%
Monterey County	88	65	73.9%
San Diego County	87	6	6.9%
Mono County	68	13	19.1%
Lake County	50	11	22.0%
Marin County	38	7	18.4%
Mendocino County	34	0	0.0%
Fresno County	32	2	6.3%
San Mateo County	31	6	19.4%
Imperial County	30	2	6.7%
San Joaquin County	28	3	10.7%
Sonoma County	24	3	12.5%
Sutter County	24	1	4.2%
Stanislaus County	21	5	23.8%
Merced County	19	5	26.3%
Solano County	16	1	6.3%
San Bernardino County	10	2	20.0%
Shasta County	10	0	0.0%

Sacramento County	9	0	0.0%
Riverside County	8	1	12.5%
Butte County	7	0	0.0%
El Dorado County	7	0	0.0%
Yuba County	7	0	0.0%
Nevada County	6	1	16.7%
Placer County	5	0	0.0%
Santa Cruz County	4	1	25.0%
Tulare County	4	1	25.0%
Yolo County	4	0	0.0%
Kings County	3	1	33.3%
Napa County	3	1	33.3%
Tehama County	3	1	33.3%
Humboldt County	2	0	0.0%
Mariposa County	2	1	50.0%
San Benito County	2	1	50.0%
Ventura County	2	1	50.0%
Amador County	1	0	0.0%
Orange County	1	0	0.0%
Santa Barbara County	1	0	0.0%
Tuolumne County	1	0	0.0%
Unknown County	7	1	14.3%

Table S3. Plasma samples from convalescent COVID-19 patients and SARS-CoV-2 vaccine recipients used for evaluating neutralizing activity against B.1.427 and B.1.429 lineage viruses

#	Sample Type	Vaccine type	Days post symptom onset	Days post-second dose	PRNT50 or 50% CPE endpoint titer		
					WA1	B.1.427 (#1-21) or B.1.429 (#22-31)	D614G
1	Convalescent Plasma	−	60	−	3200	800	NT
2	Convalescent Plasma	−	18	−	ND	ND	NT
3	Convalescent Plasma	−	70	−	3200	3200	NT
4	Convalescent Plasma	−	71	−	800	200	NT
5	Convalescent Plasma	−	85	−	400	100	NT
6	Convalescent Plasma	−	49	−	800	100	800
7	Convalescent Plasma	−	23	−	800	100	800
8	Convalescent Plasma	−	36	−	3200	1600	3200
9	Convalescent Plasma	−	31	−	3200	400	3200
10	Vaccine Recipient Plasma	BNT162b2 (Pfizer)	−	28	800	800	NT
11	Vaccine Recipient Plasma	mRNA-1273 (Moderna)	−	4	100	100	NT
12	Vaccine Recipient Plasma	BNT162b2 (Pfizer)	−	14	800	400	NT
13	Vaccine Recipient Plasma	mRNA-1273 (Moderna)	−	4	200	100	NT
14	Vaccine Recipient Plasma	mRNA-1273 (Moderna)	−	14	ND	ND	NT
15	Vaccine Recipient Plasma	mRNA-1273 (Moderna)	−	18	1600	800	NT
16	Vaccine Recipient Plasma	mRNA-1273 (Moderna)	−	10	800	400	NT
17	Vaccine Recipient Plasma	mRNA-1273 (Moderna)	−	10	3200	1600	NT

18	Vaccine Recipient Plasma	mRNA-1273 (Moderna)	−	11	3200	1600	NT
19	Vaccine Recipient Plasma	BNT162b2 (Pfizer)	−	15	200	100	400
20	Vaccine Recipient Plasma	mRNA-1273 (Moderna)	−	7	400	100	400
21	Vaccine Recipient Plasma	mRNA-1273 (Moderna)	−	5	400	400	400
22	Convalescent Plasma	−	−	55	2560	2560	2560
23	Convalescent Plasma	−	−	44	1280	640	2560
24	Convalescent Plasma	−	−	85	320	160	320
25	Convalescent Plasma	−	−	72	640	40	160
26	Convalescent Plasma	−	−	62	320	80	320
27	Convalescent Plasma	−	−	56	1280	640	2560
28	Convalescent Plasma	−	−	21	1280	640	1280
29	Convalescent Plasma	−	−	72	1280	160	640
30	Convalescent Plasma	−	−	30	1280	320	2560
31	Convalescent Plasma	−	−	42	640	80	320

Abbreviations: PRNT, plaque reduction neutralization tests; CPE, cytopathic effect; ND, antibody neutralization not detected; NT, not tested; "−", not applicable

Table S4. Sensitivity analysis exploring the effect of varying the serial interval on the calculated logistic growth rate, increase in transmission rate, and doubling time for B.1.427/B.1.429 cases in California.

Serial interval (days)	Logistic growth rate per day	Increase in transmission (%)	CI range	Doubling time (days)	CI range (doubling time)	50% date	CI width (transmission rate)
3	0.0363	10.88	9.71-12.06	19.11	17.25-21.42	1/25/21	2.35
3.5	0.0363	12.7	11.33-14.06	19.11	17.25-21.42	1/25/21	2.73
4	0.0363	14.51	12.94-16.07	19.11	17.25-21.42	1/25/21	3.13
4.5	0.0363	16.32	14.56-18.08	19.11	17.25-21.42	1/25/21	3.52
5	0.0363	18.14	16.18-20.09	19.11	17.25-21.42	1/25/21	3.91
5.5	0.0363	19.95	17.80-22.10	19.11	17.25-21.42	1/25/21	4.31
6	0.0363	21.76	19.42-24.11	19.11	17.25-21.42	1/25/21	4.69
6.5	0.0363	23.58	21.03-26.12	19.11	17.25-21.42	1/25/21	5.09
7	0.0363	25.39	22-65-28.13	19.11	17.25-21.42	1/25/21	5.48

Abbreviation: CI, 95% confidence interval

References

Avanzato, V.A., Matson, M.J., Seifert, S.N., Pryce, R., Williamson, B.N., Anzick, S.L., Barbian, K., Judson, S.D., Fischer, E.R., Martens, C.*, et al.* (2020). Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer. Cell *183*, 1901-1912 e1909.

Bedford, T., Hodcroft, E.B., and Neher, R.A. (2021). Updated Nextstrain SARS-CoV-2 clade naming strategy (NextStrain).

Bedford, T., and Neher, R. (2020). A Getting Started Guide to the Genomic Epidemiology of SARS-CoV-2. In Nextstrain documentation (Nextstrain.org).

Bushnell, B. (2021). BBMap short read aligner, and other bioinformatic tools.

Buss, L.F., Prete, C.A., Jr., Abrahim, C.M.M., Mendrone, A., Jr., Salomon, T., de Almeida-Neto, C., Franca, R.F.O., Belotti, M.C., Carvalho, M., Costa, A.G.*, et al.* (2021). Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. Science *371*, 288-292.

California (2021). Tracking COVID-19 in California (California: State of California).

Case, J.B., Rothlauf, P.W., Chen, R.E., Liu, Z., Zhao, H., Kim, A.S., Bloyet, L.M., Zeng, Q., Tahan, S., Droit, L.*, et al.* (2020). Neutralizing Antibody and Soluble ACE2 Inhibition of a Replication-Competent VSV-SARS-CoV-2 and a Clinical Isolate of SARS-CoV-2. Cell Host Microbe *28*, 475-485 e475.

CDPH (2021). COVID-19 Variant First Found in Other Countries and States Now Seen More Frequently in California (California Department of Public Health).

Challen, R., Brooks-Pollock, E., Read, J.M., Dyson, L., Tsaneva-Atanasova, K., and Danon, L. (2021). Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study. BMJ *372*, n579.

Chand, M., Hopkins, S., Dabreara, G., Allen, H., Lamagni, T., Edeghere, O., Achison, C., Myers, R., Barclay, W., Ferguson, N.*, et al.* (2020). Investigation of SARS-CoV-2 variants of concern in England, P.H. England, ed. (Crown), pp. 11.

Chen, J., Wang, R., Wang, M., and Wei, G.W. (2020). Mutations Strengthened SARS-CoV-2 Infectivity. J Mol Biol *432*, 5212-5226.

Choi, B., Choudhary, M.C., Regan, J., Sparks, J.A., Padera, R.F., Qiu, X., Solomon, I.H., Kuo, H.H., Boucau, J., Bowman, K.*, et al.* (2020). Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. N Engl J Med *383*, 2291-2293.

Cole, S., Gazy, I., Jackson, L., Hwa, S.-H., Tegally, H., Lustig, G., Giandhari, J., Pillay, S., Wilkinson, E., Naidoo, Y.*, et al.* (2021). Escape of SARS-CoV-2 501Y.V2 variants from neutralization by convalescent plasma. medRxiv.

consortiumcontact@cogconsortium.uk, C.-G.U. (2020). An integrated national scale SARS-CoV-2 genomic surveillance network. Lancet Microbe *1*, e99-e100.

Crawford, K.H.D., Eguia, R., Dingens, A.S., Loes, A.N., Malone, K.D., Wolf, C.R., Chu, H.Y., Tortorici, M.A., Veesler, D., Murphy, M.*, et al.* (2020). Protocol and Reagents for Pseudotyping Lentiviral Particles with SARS-CoV-2 Spike Protein for Neutralization Assays. Viruses *12*.

Davies, N.G., Abbott, S., Barnard, R.C., Jarvis, C.I., Kucharski, A.J., Munday, J.D., Pearson, C.A.B., Russell, T.W., Tully, D.C., Washburne, A.D.*, et al.* (2021a). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science.

Davies, N.G., Jarvis, C.I., Edmunds, W.J., Jewell, N.P., Diaz-Ordaz, K., and Keogh, R.H. (2021b). Increased hazard of death in community-tested cases of SARS-CoV-2 Variant of Concern 202012/01. medRxiv.

Davies, N.G., Jarvis, C.I., Group, C.C.-W., Edmunds, W.J., Jewell, N.P., Diaz-Ordaz, K., and Keogh, R.H. (2021c). Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. Nature.

Day, T., Gandon, S., Lion, S., and Otto, S.P. (2020). On the evolutionary epidemiology of SARS-CoV-2. Curr Biol *30*, R849-R857.

Deng, X., Gu, W., Federman, S., du Plessis, L., Pybus, O.G., Faria, N.R., Wang, C., Yu, G., Bushnell, B., Pan, C.Y.*, et al.* (2020). Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California. Science *369*, 582-587.

Drew, R.J., O'Donnell, S., LeBlanc, D., McMahon, M., and Natin, D. (2020). The importance of cycle threshold values in interpreting molecular tests for SARS-CoV-2. Diagn Microbiol Infect Dis *98*, 115130.

Drummond, A.J., Suchard, M.A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol *29*, 1969-1973.

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

Faria, N.R., Mellan, T.A., Whittaker, C., Claro, I.M., Candido, D.D.S., Mishra, S., Crispim, M.A.E., Sales, F.C., Hawryluk, I., McCrone, J.T.*, et al.* (2021). Genomics and epidemiology of a novel SARS-CoV-2 lineage in Manaus, Brazil. medRxiv.

Fontanet, A., Autran, B., Lina, B., Kieny, M.P., Karim, S.S.A., and Sridhar, D. (2021). SARS-CoV-2 variants and ending the COVID-19 pandemic. Lancet.

Gangavarapu, K., Alkuzweny, M., Cano, M., Haag, E., Latif, A.A., Mullen, J.L., Rush, B., Tsueng, G., Zhou, J., Andersen, K.G.*, et al.* (2020). outbreak.info.

Garcia-Beltran, W.F., Lam, E.C., St Denis, K., Nitido, A.D., Garcia, Z.H., Hauser, B.M., Feldman, J., Pavlovic, M.N., Gregory, D.J., Poznansky, M.C.*, et al.* (2021). Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. Cell.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A.*, et al.* (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell *181*, 271-280 e278.

Hou, Y.J., Chiba, S., Halfmann, P., Ehre, C., Kuroda, M., Dinnon, K.H., 3rd, Leist, S.R., Schafer, A., Nakajima, N., Takahashi, K.*, et al.* (2020). SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. Science *370*, 1464-1468.

Hu, J., Peng, P., Wang, K., Fang, L., Luo, F.Y., Jin, A.S., Liu, B.Z., Tang, N., and Huang, A.L. (2021). Emerging SARS-CoV-2 variants reduce neutralization sensitivity to convalescent sera and monoclonal antibodies. Cell Mol Immunol.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol *30*, 772-780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C.*, et al.* (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647-1649.

Kemp, S.A., Collier, D.A., Datir, R.P., Ferreira, I., Gayed, S., Jahun, A., Hosmillo, M., Rees-Spear, C., Mlcochova, P., Lumb, I.U.*, et al.* (2021). SARS-CoV-2 evolution during treatment of chronic infection. Nature.

Khanna, K., Raymond, W., Charbit, A.R., Jin, J., Gitlin, I., Tang, M., Sperber, H.S., Franz, S., Pillai, S., Simmons, G.*, et al.* (2020). Binding of SARS-CoV-2 spike protein to ACE2 is disabled by thiol-based drugs; evidence from in vitro SARS-CoV-2 infection studies. bioRxiv.

Klimstra, W.B., Tilston-Lunel, N.L., Nambulli, S., Boslett, J., McMillen, C.M., Gilliland, T., Dunn, M.D., Sun, C., Wheeler, S.E., Wells, A.*, et al.* (2020). SARS-CoV-2 growth, furin-cleavage-site adaptation and neutralization using serum from acutely infected hospitalized COVID-19 patients. J Gen Virol *101*, 1156-1169.

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B.*, et al.* (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell *182*, 812-827 e819.

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L.*, et al.* (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature *581*, 215-220.

Lau, E.H.Y., Tsang, O.T.Y., Hui, D.S.C., Kwan, M.Y.W., Chan, W.H., Chiu, S.S., Ko, R.L.W., Chan, K.H., Cheng, S.M.S., Perera, R.*, et al.* (2021). Neutralizing antibody titres in SARS-CoV-2 infections. Nat Commun *12*, 63.

Leung, K., Shum, M.H., Leung, G.M., Lam, T.T., and Wu, J.T. (2021). Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. Euro Surveill *26*.

Liu, Y., Liu, J., Xia, H., Zhang, X., Fontes-Garfias, C.R., Swanson, K.A., Cai, H., Sarkar, R., Chen, W., Cutler, M.*, et al.* (2021). Neutralizing Activity of BNT162b2-Elicited Serum - Preliminary Report. N Engl J Med.

Liu, Z., VanBlargan, L.A., Bloyet, L.M., Rothlauf, P.W., Chen, R.E., Stumpf, S., Zhao, H., Errico, J.M., Theel, E.S., Liebeskind, M.J.*, et al.* (2020). Landscape analysis of escape variants identifies SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. bioRxiv.

Msomi, N., Mlisana, K., de Oliveira, T., and Network for Genomic Surveillance in South Africa writing, g. (2020). A genomics network established to respond rapidly to public health threats in South Africa. Lancet Microbe *1*, e229-e230.

Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., Zhang, X., Muruato, A.E., Zou, J., Fontes-Garfias, C.R.*, et al.* (2020). Spike mutation D614G alters SARS-CoV-2 fitness. Nature.

Quick, J., Grubaugh, N.D., Pullan, S.T., Claro, I.M., Smith, A.D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T.F., Beutler, N.A.*, et al.* (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc *12*, 1261-1276.

Rai, B., Shukla, A., and Dwivedi, L.K. (2021). Estimates of serial interval for COVID-19: A systematic review and meta-analysis. Clin Epidemiol Glob Health *9*, 157-161.

Rambaut, A. (2021). FigTree.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol *67*, 901-904.

Rambaut, A., Holmes, E.C., O'Toole, A., Hill, V., McCrone, J.T., Ruis, C., du Plessis, L., and Pybus, O.G. (2020a). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol *5*, 1403-1407.

Rambaut, A., Loman, N., Pybus, O., Barclay, W., Barrett, J., Carabelli, A., Connor, T., Peacock, T., Robertson, D.L., and Volz, E. (2020b). Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. In virologicalorg, A. Network, ed.

Sabino, E.C., Buss, L.F., Carvalho, M.P.S., Prete, C.A., Jr., Crispim, M.A.E., Fraiji, N.A., Pereira, R.H.M., Parag, K.V., da Silva Peixoto, P., Kraemer, M.U.G.*, et al.* (2021). Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. Lancet *397*, 452-455.

Samuel, R.M., Majd, H., Richter, M.N., Ghazizadeh, Z., Zekavat, S.M., Navickas, A., Ramirez, J.T., Asgharian, H., Simoneau, C.R., Bonser, L.R.*, et al.* (2020). Androgen Signaling Regulates SARS-CoV-2 Receptor Levels and Is Associated with Severe COVID-19 Symptoms in Men. Cell Stem Cell *27*, 876-889 e812.

Seow, J., Graham, C., Merrick, B., Acors, S., Pickering, S., Steel, K.J.A., Hemmings, O., O'Byrne, A., Kouphou, N., Galao, R.P.*, et al.* (2020). Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans. Nat Microbiol *5*, 1598-1607.

Shu, Y., and McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill *22*.

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E.J., Msomi, N.*, et al.* (2020). Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. medRxiv, 2020.2012.2021.20248640.

Teng, S., Sobitan, A., Rhoades, R., Liu, D., and Tang, Q. (2020). Systemic effects of missense mutations on SARS-CoV-2 spike glycoprotein stability and receptor-binding affinity. Brief Bioinform.

Volz, E., Mishra, S., Chand, M., Barrett, J.C., Johnson, R., Geidelberg, L., Hinsley, W.R., Laydon, D.J., Dabrera, G., O'Toole, A.*, et al.* (2020). Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. medRxiv.

Wang, S., Sakhatskyy, P., Chou, T.H., and Lu, S. (2005). Assays for the assessment of neutralizing antibody activities against Severe Acute Respiratory Syndrome (SARS) associated coronavirus (SCV). J Immunol Methods *301*, 21-30.

Wang, Z., Schmidt, F., Weisblum, Y., Muecksch, F., Barnes, C.O., Finkin, S., Schaefer-Babajew, D., Cipolla, M., Gaebler, C., Lieberman, J.A.*, et al.* (2021). mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. Nature.

Washington, N.L., Gangavarapu, K., Zeller, M., Bolze, A., Cirulli, E.T., Schiabor Barrett, K.M., Larsen, B.B., Anderson, C., White, S., Cassens, T.*, et al.* (2021). Genomic epidemiology identifies emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. medRxiv.

Wibmer, C.K., Ayres, F., Hermanus, T., Madzivhandila, M., Kgagudi, P., Lambson, B.E., Vermeulen, M., van den Berg, K., Rossouw, T., Boswell, M.*, et al.* (2021). SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. bioRxiv.

Wise, J. (2021). Covid-19: The E484K mutation and the risks it poses. BMJ *372*.

Wu, K., Werner, A.P., Koch, M., Choi, A., Narayanan, E., Stewart-Jones, G.B.E., Colpitts, T., Bennett, H., Boyoglu-Barnum, S., Shi, W.*, et al.* (2021a). Serum Neutralizing Activity Elicited by mRNA-1273 Vaccine. N Engl J Med.

Wu, K., Werner, A.P., Moliva, J.I., Koch, M., Choi, A., Stewart-Jones, G.B.E., Bennett, H., Boyoglu-Barnum, S., Shi, W., Graham, B.S.*, et al.* (2021b). mRNA-1273 vaccine induces neutralizing antibodies against spike mutants from global SARS-CoV-2 variants. bioRxiv.

Xie, X., Liu, Y., Liu, J., Zhang, X., Zou, J., Fontes-Garfias, C.R., Xia, H., Swanson, K.A., Cutler, M., Cooper, D.*, et al.* (2021). Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. Nat Med.

Zhang, W., Davis, B.D., Chen, S.S., Sincuir Martinez, J.M., and Plummer, J.T. (2021). Emergence of a Novel SARS-CoV-2 Variant in Southern California. JAMA.

Zhou, B., Thi Nhu Thao, T., Hoffmann, D., Taddeo, A., Ebert, N., Labroussaa, F., Pohlmann, A., King, J., Steiner, S., Kelly, J.N.*, et al.* (2021). SARS-CoV-2 spike D614G change enhances replication and transmission. Nature.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R.*, et al.* (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med *382*, 727-733.

Chapter 3

Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms

David E. Gordon,# Joseph Hiatt,# Mehdi Bouhaddou,# Veronica V. Rezelj,# Svenja Ulferts,# Hannes Braberg,# Alexander S. Jureka,# Kirsten Obernier,# Jeffrey Z. Guo,# Jyoti Batra,# Robyn M. Kaake,# Andrew R. Weckstein,# Tristan W. Owens,# Meghna Gupta,# Sergei Pourmal,# Erron W. Titus#, Merve Cakir,# Margaret Soucheray, Michael McGregor, Zeynep Cakir, Gwendolyn Jang, Matthew J. O'Meara, Tia A. Tummino, Ziyang Zhang, Helene Foussard, Ajda Rojc, Yuan Zhou, Dmitry Kuchenov, Ruth Hüttenhain, Jiewei Xu, Manon Eckhardt, Danielle L. Swaney, Jacqueline M. Fabius, Manisha Ummadi, Beril Tutuncuoglu, Ujjwal Rathore, Maya Modak, Paige Haas, Kelsey M. Haas, Zun Zar Chi Naing, Ernst H. Pulido, Ying Shi, Inigo Barrio-Hernandez, Danish Memon, Eirini Petsalaki, Alistair Dunham, Miguel Correa Marrero, David Burke, Cassandra Koh, Thomas Vallet, Jesus A. Silvas, Caleigh M. Azumaya, Christian Billesbølle, Axel F. Brilot, Melody G. Campbell, Amy Diallo, Miles Sasha Dickinson, Devan Diwanji, Nadia Herrera, Nick Hoppe, Huong T. Kratochvil, Yanxin Liu, Gregory E. Merz, Michelle Moritz, Henry C. Nguyen, Carlos Nowotny, Cristina Puchades, Alexandrea N. Rizo, Ursula Schulze-Gahmen, Amber M. Smith, Ming Sun, Iris D. Young, Jianhua Zhao, Daniel Asarnow, Justin Biel, Alisa Bowen, Julian

R. Braxton, Jen Chen, Cynthia M. Chio, Un Seng Chio, Ishan Deshpande, Loan Doan, Bryan Faust, Sebastian Flores, Mingliang Jin, Kate Kim, Victor L. Lam, Fei Li, Junrui Li, Yen-Li Li, Yang Li, Xi Liu, Megan Lo, Kyle E. Lopez, Arthur A. Melo, Frank R. Moss III, Phuong Nguyen, Joana Paulino, Komal Ishwar Pawar, Jessica K. Peters, Thomas H. Pospiech Jr., Maliheh Safari, Smriti Sangwan, Kaitlin Schaefer, Paul V. Thomas, Aye C. Thwin, Raphael Trenker, Eric Tse, Tsz Kin Martin Tsui, Feng Wang, Natalie Whitis, Zanlin Yu, Kaihua Zhang, Yang Zhang, Fengbo Zhou, Daniel Saltzberg, QCRG Structural Biology Consortium†, Anthony J. Hodder, Amber S. Shun-Shion, Daniel M. Williams, Kris M. White, Romel Rosales, Thomas Kehrer, Lisa Miorin, Elena Moreno, Arvind H. Patel, Suzannah Rihn, **Mir M. Khalid**, Albert Vallejo-Gracia, Parinaz Fozouni, Camille R. Simoneau, Theodore L. Roth, David Wu, Mohd Anisul Karim, Maya Ghoussaini, Ian Dunham, Francesco Berardi, Sebastian Weigang, Maxime Chazal, Jisoo Park, James Logue, Marisa McGrath, Stuart Weston, Robert Haupt, C. James Hastie, Matthew Elliott, Fiona Brown, Kerry A. Burness, Elaine Reid, Mark Dorward, Clare Johnson, Stuart G. Wilkinson, Anna Geyer, Daniel M. Giesel, Carla Baillie, Samantha Raggett, Hannah Leech, Rachel Toth, Nicola Goodman, Kathleen C. Keough, Abigail L. Lind, Zoonomia Consortium‡, Reyna J. Klesh, Kafi R. Hemphill, Jared Carlson-Stevermer, Jennifer Oki, Kevin Holden, Travis Maures, Katherine S. Pollard, Andrej Sali, David A. Agard, Yifan Cheng, James S. Fraser, Adam Frost, Natalia Jura, Tanja Kortemme, Aashish Manglik, Daniel R. Southworth, Robert M. Stroud, Dario R. Alessi, Paul Davies, Matthew B. Frieman, Trey Ideker, Carmen Abate, Nolwenn Jouvenet, Georg Kochs, Brian Shoichet, Melanie Ott, Massimo Palmarini, Kevan M. Shokat, Adolfo García-Sastre,* Jeremy A. Rassen,* Robert Grosse,* Oren S. Rosenberg,* Kliment A. Verba,* Christopher F. Basler,* Marco Vignuzzi,* Andrew A. Peden,* Pedro Beltrao,* Nevan J. Krogan.*

Science 370, 1181 (2020)

DOI: 10.1126/science.abe9403

#These authors contributed equally *Lead contact

Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms

David E. Gordon1,2,3,4,*, Joseph Hiatt1,4,5,6,7,*, Mehdi Bouhaddou1,2,3,4,*, Veronica V. Rezelj8,*, Svenja Ulferts9,*, Hannes Braberg1,2,3,4,*, Alexander S. Jureka10,*, Kirsten Obernier1,2,3,4,*, Jeffrey Z. Guo1,2,3,4,*, Jyoti Batra1,2,3,4,*, Robyn M. Kaake1,2,3,4,*, Andrew R. Weckstein11,*, Tristan W. Owens12,*, Meghna Gupta12,*, Sergei Pourmal12,*, Erron W. Titus12,*, Merve Cakir1,2,3,4,*, Margaret Soucheray1,2,3,4, Michael McGregor1,2,3,4, Zeynep Cakir1,2,3,4, Gwendolyn Jang1,2,3,4, Matthew J. O'Meara13, Tia A. Tummino1,2,14, Ziyang Zhang1,2,3,15, Helene Foussard1,2,3,4, Ajda Rojc1,2,3,4, Yuan Zhou1,2,3,4, Dmitry Kuchenov1,2,3,4, Ruth Hüttenhain1,2,3,4, Jiewei Xu1,2,3,4, Manon Eckhardt1,2,3,4, Danielle L. Swaney1,2,3,4, Jacqueline M. Fabius1,2, Manisha Ummadi1,2,3,4, Beril Tutuncuoglu1,2,3,4, Ujjwal Rathore1,2,3,4, Maya Modak1,2,3,4, Paige Haas1,2,3,4, Kelsey M. Haas1,2,3,4, Zun Zar Chi Naing1,2,3,4, Ernst H. Pulido1,2,3,4, Ying Shi1,2,3,15, Inigo Barrio-Hernandez16, Danish Memon16, Eirini Petsalaki16, Alistair Dunham16, Miguel Correa Marrero16, David Burke16, Cassandra Koh8, Thomas Vallet8, Jesus A. Silvas10, Caleigh M. Azumaya12, Christian Billesbølle12, Axel F. Brilot12, Melody G. Campbell12,17, Amy Diallo12, Miles Sasha Dickinson12, Devan Diwanji12, Nadia Herrera12, Nick Hoppe12, Huong T. Kratochvil12, Yanxin Liu12, Gregory E. Merz12, Michelle Moritz12, Henry C. Nguyen12, Carlos Nowotny12, Cristina Puchades12, Alexandrea N. Rizo12, Ursula Schulze-Gahmen12, Amber M. Smith12, Ming Sun12,18, Iris D. Young12, Jianhua Zhao12, Daniel Asarnow12, Justin Biel12, Alisa Bowen12, Julian R. Braxton12, Jen Chen12, Cynthia M. Chio12, Un Seng Chio12, Ishan Deshpande12, Loan Doan12, Bryan Faust12, Sebastian Flores12, Mingliang Jin12, Kate Kim12, Victor L. Lam12, Fei Li12, Junrui Li12, Yen-Li Li12, Yang Li12, Xi Liu12, Megan Lo12, Kyle E. Lopez12, Arthur A. Melo12, Frank R. Moss III12, Phuong Nguyen12, Joana Paulino12, Komal Ishwar Pawar12, Jessica K. Peters12, Thomas H. Pospiech Jr.12, Maliheh Safari12, Smriti Sangwan12, Kaitlin Schaefer12, Paul V. Thomas12, Aye C. Thwin12, Raphael Trenker12, Eric Tse12, Tsz Kin Martin Tsui12, Feng Wang12, Natalie Whitis12, Zanlin Yu12, Kaihua Zhang12, Yang Zhang12, Fengbo Zhou12, Daniel Saltzberg1,2,19, QCRG Structural Biology Consortium12,†, Anthony J. Hodder20, Amber S. Shun-Shion20, Daniel M. Williams20, Kris M. White21,22, Romel Rosales21,22, Thomas Kehrer21,22, Lisa Miorin21,22, Elena Moreno21,22, Arvind H. Patel23, Suzannah Rihn23, Mir M. Khalid4, Albert Vallejo-Gracia4, Parinaz Fozouni4,5,7, Camille R. Simoneau4,7, Theodore L. Roth5,6,7, David Wu5,7, Mohd Anisul Karim24,25, Maya Ghoussaini24,25, Ian Dunham16,25, Francesco Berardi26, Sebastian Weigang27, Maxime Chazal28, Jisoo Park29, James Logue30, Marisa McGrath30, Stuart Weston30, Robert Haupt30,

C. James Hastie31, Matthew Elliott31, Fiona Brown31, Kerry A. Burness31, Elaine Reid31, Mark Dorward31, Clare Johnson31, Stuart G. Wilkinson31, Anna Geyer31, Daniel M. Giesel31, Carla Baillie31, Samantha Raggett31, Hannah Leech31, Rachel Toth31, Nicola Goodman31, Kathleen C. Keough4, Abigail L. Lind4, Zoonomia Consortium‡, Reyna J. Klesh32, Kafi R. Hemphill33, Jared Carlson-Stevermer34, Jennifer Oki34, Kevin Holden34, Travis Maures34, Katherine S. Pollard4,35,36, Andrej Sali1,2,14,19, David A. Agard1,2,12,37, Yifan Cheng1,2,12,15,37, James S. Fraser1,2,12,19, Adam Frost1,2,12,37, Natalia Jura1,2,3,12,38, Tanja Kortemme1,2,12,19,39, Aashish Manglik1,2,12,14, Daniel R. Southworth1,12,37, Robert M. Stroud1,2,12,37, Dario R. Alessi31, Paul Davies31, Matthew B. Frieman30, Trey Ideker29,40, Carmen Abate26, Nolwenn Jouvenet27,28, Georg Kochs27, Brian Shoichet1,2,14, Melanie Ott4,41, Massimo Palmarini23, Kevan M. Shokat1,2,3,15, Adolfo García-Sastre21,22,42,43,§, Jeremy A. Rassen11,§, Robert Grosse9,44,§, Oren S. Rosenberg1,2,12,36,37,41,§, Kliment A. Verba1,2,12,14,§, Christopher F. Basler10,§, Marco Vignuzzi8,§, Andrew A. Peden20,§, Pedro Beltrao16,§, Nevan J. Krogan1,2,3,4,21,§

1 Quantitative Biosciences Institute (QBI) COVID-19 Research Group (QCRG), San Francisco, CA 94158, USA.

2 QBI, University of California, San Francisco, CA 94158, USA.

3 Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA 94158, USA.

4 J. David Gladstone Institutes, San Francisco, CA 94158, USA.

5 Medical Scientist Training Program, University of California, San Francisco, CA 94143, USA.

6 Department of Microbiology and Immunology, University of California, San Francisco, CA 94143, USA.

7 Biomedical Sciences Graduate Program, University of California, San Francisco, CA 94143, USA.

8 Viral Populations and Pathogenesis Unit, CNRS UMR 3569, Institut Pasteur, 75724, Paris, cedex 15, France.

9 Institute for Clinical and Experimental Pharmacology and Toxicology I, University of Freiburg, 79104 Freiburg, Germany.

10 Center for Microbial Pathogenesis, Institute for Biomedical Sciences, Georgia State University, Atlanta, GA 30303, USA.

11 Aetion, Inc., New York, NY 10001, USA.

12 QBI Coronavirus Research Group Structural Biology Consortium, University of California, San Francisco, CA 94158, USA.

13 Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA.

14 Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94158, USA.

15 Howard Hughes Medical Institute, San Francisco, CA 94158, USA.

16 European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK.

17 Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.

18 Beam Therapeutics, Cambridge, MA 02139, USA.

19 Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA 94158, USA.

20 Department of Biomedical Science, Centre for Membrane Interactions and Dynamics, University of Sheffield, Firth Court, Sheffield S10 2TN, UK.

21 Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

22 Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

23 MRC–University of Glasgow Centre for Virus Research, Glasgow G61 1QH, Scotland, UK.

24 Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK.

25 Open Targets, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK.

26 Dipartimento di Farmacia-Scienze del Farmaco, Università degli Studi di Bari ‘ALDO MORO’, Via Orabona, 4 70125, Bari, Italy.

27 Institute of Virology, Medical Center–University of Freiburg, 79104 Freiburg, Germany.

28 Département de Virologie, CNRS UMR 3569, Institut Pasteur, Paris 75015, France.

29 Department of Medicine, University of California, San Diego, CA 92093, USA.

30 Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD 21201, USA.

31 MRC Protein Phosphorylation and Ubiquitylation Unit, College of Life Sciences, University of Dundee, Dundee DD1 5EH, UK.

32 HealthVerity, Philadelphia, PA 19103, USA.

33 Department of Neurology, University of California, San Francisco, CA 94143, USA.

34 Synthego Corporation, Redwood City, CA 94063, USA.

35 Department of Epidemiology & Biostatistics, University of California, San Francisco, CA 94158, USA.

36 Chan-Zuckerberg Biohub, San Francisco, CA 94158, USA.

37 Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94158, USA.

38 Cardiovascular Research Institute, University of California, San Francisco, CA 94158, USA.

39 The University of California, Berkeley–University of California, San Francisco Graduate Program in Bioengineering, University of California, San Francisco, CA 94158, USA.

40 Department to Bioengineering, University of California, San Diego, CA 92093, USA.

41 Department of Medicine, University of California, San Francisco, CA 94143, USA.

42 Department of Medicine, Division of Infectious Diseases, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

43 The Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

44 Centre for Integrative Biological Signaling Studies (CIBSS), University of Freiburg, 79104 Freiburg, Germany.

§ Corresponding author. Email: nevan.krogan@ucsf.edu (N.J.K.); pbeltrao@ebi.ac.uk (P.B.); marco.vignuzzi@pasteur.fr (M.V.); cbasler@gsu.edu (C.F.B.); verba@msg.ucsf.edu (K.A.V.); oren.rosenberg@ucsf.edu (O.S.R.); a.peden@sheffield.ac.uk (A.A.P.); robert.grosse@pharmakol.uni-freiburg.de (R.G.); jeremy.rassen@aetion.com (J.A.R.); Adolfo.Garcia-Sastre@mssm.edu (A.G.-S.)

* These authors contributed equally to this work.

† The QCRG Structural Biology Consortium collaborators and their affiliations are listed in the supplementary materials.

‡ The Zoonomia Consortium collaborators and their affiliations are listed in the supplementary materials.

Abstract

The COVID-19 pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is a grave threat to public health and the global economy. SARS-CoV-2 is closely related to the more lethal but less transmissible coronaviruses SARS-CoV-1 and Middle East respiratory syndrome coronavirus (MERS-CoV). Here, we have carried out comparative viral-human protein-protein interaction and viral protein localization analyses for all three viruses. Subsequent functional genetic screening identified host factors that functionally impinge on coronavirus proliferation, including Tom70, a mitochondrial chaperone protein that interacts with both SARS-CoV-1 and SARS-CoV-2 ORF9b, an interaction we structurally characterized using cryo–electron microscopy. Combining genetically validated host factors with both COVID-19 patient genetic data and medical billing records identified molecular mechanisms and potential drug treatments that merit further molecular and clinical study.

Introduction

In the past two decades, three deadly human respiratory syndromes associated with coronavirus (CoV) infections have emerged: severe acute respiratory syndrome (SARS) in 2002, Middle East respiratory syndrome (MERS) in 2012, and COVID-19 in 2019. These three diseases are caused by the zoonotic coronaviruses severe acute respiratory syndrome coronavirus 1 (SARS-CoV-1), Middle East respiratory syndrome coronavirus (MERS-CoV), and SARS-CoV-2 (1), respectively. Before their emergence, human coronaviruses were associated with usually mild respiratory illness. To date, SARS-CoV-2 has sickened millions and killed more than 1 million people worldwide. This unprecedented challenge has prompted widespread efforts to develop vaccine and antiviral strategies, including repurposed therapeutics, which offer the potential for treatments with known safety profiles and short development timelines. The successful repurposing of the antiviral nucleoside analog Remdesivir (2) as well as the host-directed anti-inflammatory steroid dexamethasone (3) provide clear proof that existing compounds can be crucial tools in the fight against COVID-19. Despite these promising examples, there is still no curative treatment for COVID-19. Additionally, as with any virus, the search for effective antiviral strategies could be complicated over time by the continued evolution of SARS-CoV-2 and possible resulting drug resistance (4).

Current endeavors are appropriately focused on SARS-CoV-2 because of the severity and urgency of the ongoing pandemic. However, the frequency with which highly virulent coronavirus strains have emerged highlights an additional need to identify promising targets for broad coronavirus inhibitors with high barriers to resistance mutations and the potential for rapid deployment against future emerging strains. Although traditional antivirals target viral enzymes that are often subject to mutation and thus the development of drug resistance, targeting the host proteins required for viral replication is a strategy that can avoid resistance and lead to therapeutics with the potential for broad-spectrum activity because families of viruses often exploit common cellular pathways and processes.

Here, we identified shared biology and potential drug targets among the three highly pathogenic human coronavirus strains. We expanded on our recently published map of virus-host protein interactions for SARS-CoV-2 (5) and mapped the full interactomes of SARS-CoV-1 and MERS-CoV. We investigated the localization of viral proteins across strains and quantitatively compared the virus-human interactions for each virus. Using functional genetics and structural analysis of selected host-dependency factors, we identified drug targets and performed real-world analysis on clinical data from COVID-19 patients.

A cross-coronavirus study of protein function

A central goal of this study is to understand, from a systems level, the conservation of target proteins and cellular processes between SARS-CoV-2, SARS-CoV-1, and MERS-CoV, thereby identifying shared vulnerabilities that can be targeted with antiviral therapeutics. All three strains encode four homologous structural proteins (E, M, N, and S) and 16 nonstructural proteins (Nsps). The latter are proteolytically cleaved from a polyprotein precursor that is expressed from one large open reading frame (ORF), ORF1ab (Fig. 1A). Additionally, coronaviruses contain a variable number of accessory factors encoded by ORFs. Although the genome organization and sequence of ORF1ab is mainly conserved between the three viruses under study, it diverges markedly in the region encoding the accessory factors, especially between MERS-CoV and the two SARS coronaviruses (Fig. 1, A to D, and table S1). These differences in conservation of genes and genome organization are linked to differences in host-targeting systems that we have studied through large-scale protein localization and interaction profiling (Fig. 1E). Building on our earlier work on the interactome of SARS-CoV-2 (5), we identified the host factors physically interacting with each SARS-CoV-1 and MERS-CoV viral protein. To this end, structural proteins, mature Nsps, and predicted ORF proteins were codon optimized, 2xStrep tagged, and cloned into a mammalian expression vector (figs. S1 and S2; see below and Materials and methods section). Each protein construct was transfected into HEK293T cells and affinity purified, and high-confidence interactors were identified by mass spectrometry (MS) and scored using SAINTexpress (significance analysis of interactome) and MiST (mass spectrometry interaction statistics) scoring algorithms (6, 7) (table S2 and figs. S3 to S6). Additionally, we performed MS analysis on SARS-CoV-2 Nsp16, which was not analyzed in our earlier work (5) (table S2 and fig. S7). In all, we now report 389 high-confidence interactors for SARS-CoV-2, 366 interactors for SARS-CoV-1, and 296 interactors for MERS-CoV (table S2).

Conserved coronavirus proteins often retain the same cellular localization

As protein localization can provide important information regarding function, we assessed the cellular localization of individually expressed coronavirus proteins in addition to mapping their interactions (Fig. 2A and Materials and methods). Immunofluorescence localization analysis of all 2xStrep-tagged SARS-CoV-2, SARS-CoV-1, and MERS-CoV proteins highlights similar patterns of localization for most shared protein homologs in HeLaM cells (Fig. 2B), which supports the hypothesis that conserved proteins share functional similarities. A notable exception is Nsp13, which appears to localize to the cytoplasm for SARS-CoV-2 and SARS-CoV-1, but to the mitochondria for MERS-CoV (Fig. 2B, figs. S8 to S13, and table S3). To assess the localization

of SARS-CoV-2 proteins in the context of infected cells, we raised antibodies against 20 SARS-CoV-2 proteins and validated them with the individually expressed 2xStrep-tagged proteins (fig. S14). Using the 14 antibodies with confirmed specificity, we observed that localization of viral proteins in infected Caco-2 cells sometimes differed from their localization when expressed individually (Fig. 2B, fig. S15, and table S3). This likely results from recruitment of viral proteins and complexes into replication compartments, as well as from remodeling of the secretory pathway during viral infection. Such differences could also result from mislocalization caused by protein tagging. For example, the localization of expressed ORF7B does not match the known SARS-CoV-1 Golgi localization seen in the infection state. For proteins such as Nsp1 and ORF3a, which are not known to be involved in viral replication, their localization is consistent both when expressed individually and in the context of viral infection (Fig. 2, C and D). We have compared the localization of the expressed viral proteins with the localization of their interaction partners using a cellular compartment gene ontology (GO) enrichment analysis (fig. S16). Several examples exist where the localization of the viral protein is in agreement with the localization of the interaction partners, including enrichment of the nuclear pore for Nsp9 interactors and endoplasmic reticulum (ER) enrichment for interactions with ORF8.

Our localization studies suggest that most orthologous proteins have the same localization across the viruses (Fig. 2B). Moreover, small changes in localization, as observed for some viral proteins across strains, do not coincide with strong changes in virus-host protein interactions (Fig. 2E). Overall, these results suggest that changes in protein localization, as measured by expressed, tagged proteins, are not common and therefore are unlikely to be a major source of differences in host-targeting mechanisms.

Comparison of host-targeted processes identifies conserved mechanisms with divergent implementations

To study the conservation of targeted host factors and processes, we first used a clustering approach (Materials and methods) to compare the overlap in protein interactions for the three viruses (Fig. 3A). We defined seven clusters of virus-host interactions corresponding to those that are specific to each virus or are shared among sets of viruses. The largest pairwise overlap was observed between SARS-CoV-1 and SARS-CoV-2 (Fig. 3A), as is expected from their closer evolutionary relationship. A functional enrichment analysis (Fig. 3B and table S4) highlighted host processes that are targeted through interactions conserved across all three viruses, including ribosome biogenesis and regulation of RNA metabolism. Conserved interactions between SARS-CoV-1 and SARS-CoV-2—but not MERS-CoV—were enriched in endosomal and Golgi vesicle

transport (Fig. 3B). Despite the small fraction (7.1%) of interactions conserved between SARS-CoV-1 and MERS-CoV—but not SARS-CoV-2—these were strongly enriched in translation initiation and myosin complex proteins (Fig. 3B).

We next asked whether the conserved interactions were specific for certain viral proteins (Fig. 3C) and found that some proteins (M, N, Nsp7, Nsp8, and Nsp13) showed a disproportionately high fraction of shared interactions conserved across the three viruses. This suggests that the processes targeted by these proteins may be more essential and more likely to be required for other emerging coronaviruses. Such differences in conservation of interactions should be encoded, to some extent, in the degree of sequence differences. Comparing pairs of homologous proteins shared between SARS-CoV-2 and SARS-CoV-1 or MERS-CoV, we observed a significant correlation between sequence conservation and protein-protein interaction (PPI) similarity (calculated as Jaccard index) [Fig. 3D; correlation coefficient (r) = 0.58, P = 0.0001]. This shows that the evolution of protein sequences strongly determines the divergence in virus-host interactions.

While studying the function of host proteins interacting with each virus, we noted that some shared cellular processes were targeted by different interactions across the viruses. To study this in more detail, we identified the cellular processes significantly enriched in the interactomes of all three viruses (fig. S17A and table S4) and ranked them by the degree of overlapping proteins (Fig. 3E). This identified proteins related to the nuclear envelope, proteasomal catabolism, cellular response to heat, and regulation of intracellular protein transport as biological functions that are hijacked by these viruses through different human proteins. Additionally, we found that up to 51% of protein interactions with a conserved human target occurred via a different (nonorthologous) viral protein (Fig. 3F), and, in some cases, the overlap of interactions for two nonorthologous virus baits was greater than that for the orthologous pair (Fig. 3G and fig. S17, B and C). For example, several interacting proteins of SARS-CoV-2 Nsp8 are also targeted by MERS-CoV ORF4a, and interactions of MERS-CoV ORF5 share interactors with SARS-CoV-2 ORF3a (Fig. 3G). In the case of Nsp8, we found some degree of structural homology between its C-terminal region and a predicted structural model of ORF4a (Materials and methods and fig. S17D), which is indicative of a possible common interaction mechanism.

We find that sequence differences determine the degree of changes in virus-host interactions and that often the same cellular process can be targeted by different viral or host proteins. These results suggest a degree of plasticity in the way that these viruses can control a given biological process in the host cell.

Quantitative differential interaction scoring identifies interactions conserved between coronaviruses

The identification of virus-host interactions conserved across pathogenic coronaviruses provides the opportunity to reveal host targets that may remain essential for these and other emerging coronaviruses. For a quantitative comparison of each virus-human interaction from viral baits shared by all three viruses, we developed a differential interaction score (DIS). A DIS is calculated between any pair of viruses and is defined as the difference between the interaction scores (K) from each virus (Fig. 4A, table S5, and Materials and methods). This kind of comparative analysis is beneficial as it permits the recovery of conserved interactions that may fall just below strict cutoffs. For each comparison, a DIS was calculated for interactions residing in certain clusters as defined in the previous analysis (see Fig. 3A). For example, for the SARS-CoV-2 to MERS-CoV comparison, a DIS was computed for interactions residing in all clusters except cluster 3, where interactions are either not found or scores were very low for both SARS-CoV-2 and MERS-CoV. A DIS of 0 indicates that the interaction is confidently shared between the two viruses being compared, whereas a DIS of +1 or −1 indicates that the host-protein interaction is specific for the virus listed first or second, respectively.

In agreement with our previous results (Fig. 3A), DIS values for the comparison between SARS-CoV-2 and SARS-CoV-1 are enriched near zero, which indicates a high number of shared interactions (Fig. 4B, yellow). By contrast, comparing interactions from either SARS-CoV-1 or SARS-CoV-2 with MERS-CoV resulted in DIS values closer to ±1, which indicates a higher divergence (Fig. 4B, blue and green). The breakdown of DISs by homologous viral proteins reveals a high similarity of interactions for proteins N, Nsp8, Nsp7, and Nsp13 (Fig. 4C), reinforcing the observations made by overlapping thresholded interactions (Fig. 3, C and D). As the greatest dissimilarity was observed between the SARS coronaviruses and MERS-CoV, we computed a fourth DIS (SARS-MERS) by averaging K from SARS-CoV-1 and SARS-CoV-2 before calculating the difference with MERS-CoV (Fig. 4, B and C, purple). We next created a network visualization of the SARS-MERS comparison (Fig. 4D), permitting an appreciation of SARS-specific (red; DIS near +1) versus MERS-specific (blue; DIS near −1) interactions as well as those conserved between all three coronavirus species (black; DIS near 0). SARS-specific interactions include DNA polymerase α interacting with Nsp1, stress granule regulators interacting with N protein, TLE transcription factors interacting with Nsp13, and AP2 clathrin interacting with Nsp10. Notable MERS-CoV–specific interactions include mammalian target of rapamycin (mTOR) and Stat3 interacting with Nsp1; DNA damage response components p53 (TP53),

MRE11, RAD50, and UBR5 interacting with Nsp14; and the activating signal cointegrator 1 (ASC-1) complex interacting with Nsp2. Interactions shared between all three coronaviruses include casein kinase II and RNA processing regulators interacting with N protein; inosine 5′-monophosphate (IMP) dehydrogenase 2 (IMPDH2) interacting with Nsp14; centrosome, protein kinase A, and TBK1 interacting with Nsp13; and the signal recognition particle, 7SK small nuclear ribonucleoprotein (snRNP), exosome, and ribosome biogenesis components interacting with Nsp8 (Fig. 4D).

Cell-based genetic screens identify SARS-CoV-2 host-dependency factors

To identify host factors that are critical for infection and therefore potential targets for host-directed therapies, we performed genetic perturbations of 332 human proteins—331 previously identified to interact with SARS-CoV-2 proteins (5) plus ACE2—and observed their effect on infectivity. To ensure a broad coverage of potential hits, we carried out two screens in different cell lines, investigating the effects on infection: small interfering RNA (siRNA) knockdowns in A549 cells stably expressing ACE2 (A549-ACE2) (Fig. 5A) and CRISPR-based knockouts in Caco-2 cells (Fig. 5B). ACE2 was included as positive control in both screens as were nontargeting siRNAs or nontargeted Caco-2 cells as negative controls. After SARS-CoV-2 infection, effects on virus infectivity were quantified by real-time quantitative polymerase chain reaction (RT-qPCR) on cell supernatants (siRNA) or by titrating virus-containing supernatants on Vero E6 cells (CRISPR) (see Materials and methods for details). Cells were monitored for viability, and knockdown or editing efficiency was determined as described (Materials and methods and fig. S18). This revealed that 93% of the genes were knocked down at least 50% in the A549-ACE2 screen, and 95% of the knockdowns exhibited a <20% decrease in viability. In the Caco-2 assay, we observed an editing efficiency of at least 80% for 89% of the genes tested (Materials and methods and fig. S18). Of the 332 human SARS-CoV-2 interactors, the final A549-ACE2 dataset includes 331 gene knockdowns, and the Caco-2 dataset includes 286 gene knockouts, with the difference mainly owing to the removal of essential genes (Materials and methods). The readouts from both assays were then separately normalized using robust z-scores (Materials and methods), with negative and positive z-scores indicating proviral-dependency factors (perturbation leads to decreased infectivity) and antiviral host factors with restrictive activity (perturbation leads to increased infectivity), respectively. As expected, negative controls resulted in neutral z-scores (Fig. 5, C and D, and tables S6 and S7). Similarly, perturbations of the positive control ACE2 resulted in strongly negative z-scores in both assays (Fig. 5, C and D). Overall, the z-scores did not exhibit any trends related to viability, knockdown efficiency, or editing efficiency (fig. S18). With a cutoff of $|z|>2$ to

highlight genes that notably affect SARS-CoV-2 infectivity when perturbed, 31 and 40 dependency factors ($z < -2$) and 3 and 4 factors with restrictive activity ($z > 2$) were identified in A549-ACE2 and Caco-2 cells, respectively (Fig. 5E). Of particular interest are the host-dependency factors for SARS-CoV-2 infection, which represent potential targets for drug development and repurposing. For example, nonopioid receptor sigma 1 (sigma-1, encoded by SIGMAR1) was identified as a functional host-dependency factor in both cell systems, in agreement with our previous report of antiviral activity for sigma receptor ligands (5). To provide a contextual view of the genetics results, we generated a network that integrates the hits from both cell lines and the PPIs of their encoded proteins with SARS-CoV-2, SARS-CoV-1, and MERS-CoV proteins (Fig. 5F). Notably, we observed an enrichment of genetic hits that encode proteins interacting with viral Nsp7, which has a high degree of interactions shared across all three viruses (Fig. 3C). Prostaglandin E synthase 2 (PGES2, encoded by PTGES2), for example, is a functional interactor of Nsp7 from SARS-CoV-1, SARS-CoV-2, and MERS-CoV. Other dependency factors were specific to SARS-CoV-2, including interleukin-17 (IL-17) receptor A (IL17RA), which interacts with SARS-CoV-2 ORF8. We also identify dependency factors that are shared interactors between SARS-CoV-1 and SARS-CoV-2 such as the aforementioned sigma receptor 1 (SIGMAR1), which interacts with Nsp6, and the mitochondrial import receptor subunit Tom70 (TOMM70), which interacts with ORF9b. We will use these interactions to validate virus-host interactions (ORF8-IL17RA and ORF9b-Tom70), connect our systems biology data to evidence for the clinical impact of the host factors we identified (IL17RA), and analyze outcomes of COVID-19 patients treated with putative host-directed drugs against PGES-2 and sigma receptor 1.

SARS ORF9b interacts with Tom70

ORF9b of SARS-CoV-1 and SARS-CoV-2 were found to be localized to mitochondria upon overexpression as well as in SARS-CoV-2–infected cells. In line with this, the mitochondrial outer membrane protein Tom70 (encoded by TOMM70) is a high-confidence interactor of ORF9b in both SARS-CoV-1 and SARS-CoV-2 interaction maps (Fig. 6A), and it acts as a host-dependency factor for SARS-CoV-2 (Fig. 6B). Tom70 falls below the scoring threshold as a putative interactor of MERS-CoV Nsp2, a viral protein not associated with mitochondria (table S2). Tom70 is one of the major import receptors in the translocase of the outer membrane (TOM) complex that recognizes and mediates the translocation of mitochondrial preproteins from the cytosol into the mitochondria in a chaperone-dependent manner (8). Additionally, Tom70 is involved in the

activation of the mitochondrial antiviral signaling (MAVS) protein, which leads to apoptosis upon viral infection (9, 10).

To validate the interaction between viral proteins and Tom70, we performed a coimmunoprecipitation experiment in the presence or absence of Strep-tagged ORF9b from SARS-CoV-1 and SARS-CoV-2 as well as Strep-tagged Nsp2 from all three coronaviruses. Endogenous Tom70—but not other translocase proteins of the outer membrane including Tom20, Tom22, and Tom40—coprecipitated only in the presence of ORF9b but not Nsp2 in both HEK293T and A549 cells, which confirms our affinity purification–mass spectrometry (AP-MS) data and suggests that ORF9b specifically interacts with Tom70 (Fig. 6C and fig. S19A). Further, upon coexpression in bacterial cells, we were able to copurify the ORF9b-Tom70 protein complex, which indicates a stable complex (Fig. 6D). We found that SARS-CoV-1 and SARS-CoV-2 ORF9b expressed in HeLaM cells colocalized with Tom70 (Fig. 6E) and observed that SARS-CoV-1 or SARS-CoV-2 ORF9b overexpression led to decreases in Tom70 expression (Fig. 6, E and F). Similarly, ORF9b was found to colocalize with Tom70 on SARS-CoV-2 infection (Fig. 6G). This is in agreement with the known outer mitochondrial membrane localization of Tom70 (11) and ORF9b localization to mitochondria upon overexpression and during SARS-CoV-2 infection (Fig. 2B). We also saw decreases in Tom70 expression during SARS-CoV-2 infection (Fig. 6G) but did not see pronounced changes in expression levels of the mitochondrial protein Tom20 after individual Strep-ORF9b expression or upon SARS-CoV-2 infection (fig. S19, B and C).

Cryo–electron microscopy structure of ORF9b-Tom70 complex reveals ORF9b interacting at the substrate binding site of Tom70

Tom70, as part of the TOM complex, is involved in the recognition of mitochondrial preproteins from the cytosol (12). To further understand the molecular details of ORF9b-Tom70 interactions, we obtained a 3-Å cryo–electron microscopy (cryo-EM) structure of the ORF9b-Tom70 complex (Fig. 7A and fig. S20). Notably, although purified proteins failed to interact upon attempted in vitro complex reconstitution, they yielded a stable and pure complex when coexpressed in Escherichia coli (Fig. 6D). This may be because of the fact that ORF9b alone purifies as a dimer (as inferred by the apparent molecular weight on size-exclusion chromatography) and would need to dissociate to interact with Tom70 on the basis of our structure. Tom70 preferentially binds preproteins with internal hydrophobic targeting sequences (13). It contains an N-terminal transmembrane domain and tetratricopeptide repeat (TPR) motifs in its cytosolic segment. The C-terminal TPR motifs recognize the internal mitochondrial targeting signals (MTSs) of preproteins, and the N-terminal TPR clamp domain serves as a docking site for multichaperone

complexes that contain preprotein (14, 15). Obtained cryo-EM density allowed us to build atomic models for residues 109 to 600 of human Tom70 and residues 39 to 76 of SARS-CoV-2 ORF9b (Fig. 7A and table S8). ORF9b makes extensive hydrophobic interactions at the pocket on Tom70 that have been implicated in its binding to MTS, with the total buried surface area at the interface being quite extensive—~2000 Å2 (Fig. 7B). In addition to the mostly hydrophobic interface, four salt bridges further stabilize the interaction (Fig. 7C). On interaction with ORF9b, the interacting helices on Tom70 move inward to tightly wrap around ORF9b as compared with previously crystallized yeast Tom70 homologs (movie S1). No structure for human Tom70 without a substrate has been reported to date, and therefore we cannot rule out the idea that the conformational differences are because of differences between homologs. However, it is possible that this conformational change upon substrate binding is conserved across homologs, as many of the Tom70 residues interacting with ORF9b are highly conserved, which likely indicates residues essential for endogenous MTS substrate recognition.

Although a previously published crystal structure of SARS-CoV-2 ORF9b revealed that it entirely consists of beta sheets [Protein Data Bank (PDB) ID: 6Z4U] (16), we observed that, upon binding Tom70 residues 52 to 68, ORF9b forms a helix (Fig. 7D). This is consistent with the fact that MTS sequences recognized by Tom70 are usually helical, and analysis with the TargetP MTS prediction server revealed a high probability for this region of ORF9b to have an MTS (Fig. 7E). This shows structural plasticity in this viral protein where, depending on the binding partner, ORF9b changes between helical and beta strand folds. Furthermore, we had previously identified two infection-driven phosphorylation sites on ORF9b, S50 and S53 (17), which map to the region on ORF9b buried deep in the Tom70 binding pocket (Fig. 7B, yellow). S53 contributes two hydrogen bonds to the interaction with Tom70 in this overall hydrophobic region. Therefore, once phosphorylated, it is likely that the ORF9b-Tom70 interaction is weakened. These residues are surface exposed in the dimeric structure of ORF9b, which could potentially allow phosphorylation to partition ORF9b between Tom70-bound and dimeric populations.

The two binding sites on Tom70—the substrate binding site and the TPR domain that recognizes Hsp70 and Hsp90—are known to be conformationally coupled (18). Tom70's interaction with a C-terminal EEVD motif of Hsp90 via the TPR domain is key for its function in the interferon pathway and induction of apoptosis on viral infection (10, 19). Whether ORF9b, by binding to the substrate recognition site of Tom70, allosterically inhibits Tom70's interaction with Hsp90 at the TPR domain remains to be investigated; but notably we observe in our structure that R192, a key

residue in the interaction with Hsp70 and Hsp90, is moved out of position to interact with the EEVD sequence, which suggests that ORF9b may modulate interferon and apoptosis signaling via Tom70 (fig. S21). Alternatively, Tom70 has been described as an essential import receptor for PTEN induced kinase 1 (PINK1), and therefore the loss of mitochondrial import efficiency as a result of ORF9b binding to the Tom70 substrate binding pocket may induce mitophagy.

Implications of the ORF8-IL17RA interaction for COVID-19

As described above, we found that IL-17 receptor A (IL17RA) physically interacts with ORF8 from SARS-CoV-2, but not SARS-CoV-1 or MERS-CoV (Fig. 5D, table S2, and Fig. 8A). Several recent studies have identified high IL-17 levels or aberrant IL-17 signaling as a correlate of severe COVID-19 (20–23). We demonstrated that the physical interaction of SARS-CoV-2 ORF8 with IL17RA occurs with or without IL-17A treatment, which suggests that signaling through the receptor does not disrupt the interaction with ORF8 (Fig. 8B). Furthermore, knockdown of IL17RA led to a significant decrease in SARS-CoV-2 viral replication in A549-ACE2 cells (Fig. 8C). These data suggest that the ORF8-IL17RA interaction modulates systemic IL-17 signaling.

One manner in which this signaling is regulated is through the release of the extracellular domain of the receptor as soluble IL17RA (sIL17RA), which acts as a decoy in circulation by soaking up IL-17A and inhibiting IL-17 signaling (24). Production of sIL17RA has been demonstrated by alternative splicing in cultured cells (25), but the mechanism by which IL17RA is shed in vivo remains unclear (26). ADAM family metalloproteases are known to mediate the release of other interleukin receptors into their soluble form (27). We found that SARS-CoV-2 ORF8 physically interacted with both ADAM9 and ADAMTS1 in our previous study (5). We find that knockdown of ADAM9, like that of IL17RA, leads to significant decreases in SARS-CoV-2 replication in A549-ACE2 cells (Fig. 5D and table S2).

To test the in vivo relevance of sIL17RA in modulating SARS-CoV-2 infection, we leveraged a genome-wide association study (GWAS) which identified 14 single-nucleotide polymorphisms (SNPs) near the IL17RA gene that causally regulate sIL17RA plasma levels (28). We then used generalized summary-based Mendelian randomization (GSMR) (28, 29) on the curated GWAS datasets of the COVID-19 Host Genetics Initiative (COVID-HGI) (30) and observed that genotypes that predicted higher sIL17RA plasma levels were associated with lower risk of COVID-19 when compared with the population (Fig. 8D and table S9), which is seemingly consistent with our molecular data. Similar results were obtained when comparing only hospitalized COVID-19 patients to the population. However, there was no evidence of association in hospitalized versus

nonhospitalized COVID-19 patients. Though the COVID-HGI dataset is underpowered and this observation needs to be replicated in other cohorts, the clinical observations, functional genetics, and clinical genetics all suggest that SARS-CoV-2 benefits from modulating IL-17 signaling. One potentially contradictory caveat is that we find high-level IL-17A treatment diminishes SARS-CoV-2 replication in A549-ACE2 cells (fig. S22); however, IL-17 is a pleiotropic cytokine and it is likely to play multiple roles during SARS-CoV-2 infection in the context of a competent immune system.

Infectious and transmissible SARS-CoV-2 viruses with large deletions of ORF8 have arisen during the pandemic and have been associated with milder disease and lower concentrations of proinflammatory cytokines (31). Notably, compared with healthy controls, patients infected with wildtype, but not ORF8-deleted virus, had threefold elevated plasma levels of IL-17A (31). More work will be needed to understand if and how ORF8 manipulates the IL-17 signaling pathway during the course of SARS-CoV-2 infection.

Investigation of druggable targets identified as interactors of multiple coronaviruses

The identification of druggable host factors provides a rationale for drug repurposing efforts. Given the extent of the current pandemic, real-world data can now be used to study the outcome of COVID-19 patients coincidentally treated with host factor–directed, U.S. Food and Drug Administration (FDA)–approved therapeutics. Using medical billing data, we identified 738,933 patients in the United States with documented SARS-CoV-2 infection (Materials and methods). In this cohort, we probed the use of drugs against targets identified here that were shared across coronavirus strains and found to be functionally relevant in the genetic perturbation screens. In particular, we analyzed outcomes for an inhibitor of prostaglandin E synthase type 2 (PGES-2, encoded by PTGES2) and for potential ligands of sigma nonopioid receptor 1 (sigma-1, encoded by SIGMAR1), and investigated whether these patients fared better than carefully matched patients treated with clinically similar drugs without predicted anticoronavirus activity.

PGES-2, an interactor of Nsp7 from all three viruses (Fig. 4D), is a dependency factor for SARS-CoV-2 (Fig. 5F). It is inhibited by the FDA-approved prescription nonsteroidal anti-inflammatory drug (NSAID) indomethacin. Computational docking of Nsp7 and PGES-2 to predict binding configuration showed that the dominant cluster of models localizes Nsp7 adjacent to the PGES-2–indomethacin binding site (fig. S23). However, indomethacin did not inhibit SARS-CoV-2 in vitro at reasonable antiviral concentrations (fig. S24 and table S10). A previous study also found that similarly high levels of the drug were needed for inhibition of SARS-CoV-1 in vitro, but this study still showed efficacy for indomethacin against canine coronavirus in vivo (32). This motivated us

to observe outcomes in a cohort of outpatients with confirmed SARS-CoV-2 infection who by happenstance initiated a course of indomethacin compared with those who initiated the prescription NSAID celecoxib, which lacks anti–PGES-2 activity. We compared the odds of hospitalization by risk-set sampling (RSS) patients treated at the same time and at similar levels of disease severity and then by further matching on propensity score (PS) (33) (Fig. 9A and table S11). RSS and PS—combined with a new user, active comparator design that mimics the interventional component of parallel group randomized studies—are established design and analytic techniques that mitigate biases that can arise in observational studies. A complete list of risk factors used for matching, which include demographic data, baseline health care utilization, comorbidities, and measures of disease severity, are found in table S11.

Among SARS-CoV-2–positive patients, new users of indomethacin in the outpatient setting were less likely than matched new users of celecoxib to require hospitalization or inpatient services [Fig. 9B; odds ratio (OR) = 0.33; 95% confidence interval (CI): 0.03 to 3.19]. The CI of our primary analysis included the null value. In sensitivity analyses, neither using the larger, risk-set–sampled cohort nor relaxing our outcome definition to include any hospital visit appreciably changed the interpretation of our findings, but it did narrow the CIs, particularly when both approaches were combined (OR = 0.25; 95% CI: 0.08 to 0.76). Although we acknowledge that this is a small, noninterventional study, it is nonetheless an example of how molecular insight can rapidly generate testable clinical hypotheses and help prioritize candidates for prospective clinical trials or future drug development.

To create larger patient cohorts, we next grouped drugs that shared activity against the same target—sigma receptors. We previously identified sigma-1 and sigma-2 as drug targets in our SARS-CoV-2–human PPI map, and multiple potent, nonselective sigma ligands were among the most promising inhibitors of SARS-CoV-2 replication in Vero E6 cells (5). As shown above, knockout and knockdown of SIGMAR1, but not of SIGMAR2 (also known as TMEM97), led to robust decreases in SARS-CoV-2 replication (fig. S24 and Fig. 5F), which suggests that sigma-1 may be a key therapeutic target. We analyzed SIGMAR1 sequences across 359 mammals and observed positive selection of several residues within beaked whale, mouse, and ruminant lineages, which may indicate a role in host-pathogen competition (fig. S25). Additionally, the sigma ligand drug amiodarone inhibited replication of SARS-CoV-1 as well as SARS-CoV-2, consistent with the conservation of the Nsp6–sigma-1 interaction across the SARS viruses (fig. S24 and Fig. 4D). We then looked for other FDA-approved drugs with reported nanomolar affinity for sigma receptors or those that fit the sigma ligand chemotype (5, 34–41), and we selected 13

such therapeutics. We find that all are potent inhibitors of SARS-CoV-2 with half-maximal inhibitory concentration (IC50) values <10 μM, though there is a wide range in reported sigma receptor affinity with no clear correlation between sigma receptor binding affinity and antiviral activity (fig. S24D). Several clinical drug classes were represented by more than one candidate, including typical antipsychotics and antihistamines. Over-the-counter antihistamines are not well represented in medical billing data and are therefore poor candidates for real-world analysis, but users of typical antipsychotics can be easily identified in our patient cohort. By grouping these individual drug candidates by clinical indication, we were able to build a better-powered comparison.

We constructed a cohort for retrospective analysis on new, inpatient users of antipsychotics. In inpatient settings, typical and atypical antipsychotics are used similarly, most commonly for delirium. We compared the effectiveness of typical antipsychotics, which have sigma activity and antiviral effects (fig. S24E), versus atypical antipsychotics, which do not have antiviral activity (fig. S24F), for treatment of COVID-19 (Fig. 9C). Observing mechanical ventilation outcomes in inpatient cohorts is a proxy for the worsening of severe illness rather than the progression from mild disease signified by the hospitalization of indomethacin-exposed outpatients above. We again used RSS plus PS to build a robust, directly comparable cohort of inpatients (table S11). In our primary analysis, half as many of the new users of typical antipsychotics compared with the new users of atypical antipsychotics progressed to the point of requiring mechanical ventilation, demonstrating significantly lower use with an OR of 0.46 (95% CI: 0.23 to 0.93; P = 0.03; Fig. 9D). As above, we conducted a sensitivity analysis in the RSS-only cohort and observed the same trend (OR = 0.56; 95% CI: 0.31 to 1.02; P = 0.06), which emphasizes the primary result of a beneficial effect for typical versus atypical antipsychotics observed in the RSS-plus-PS–matched cohort. Although a careful analysis of the relative benefits and risks of typical antipsychotics should be undertaken before considering prospective studies or interventions, these data and analyses demonstrate how molecular information can be translated into real-world implications for the treatment of COVID-19—an approach that can ultimately be applied to other diseases in the future.

Discussion

In this study, we generated and compared three different coronavirus-human PPI maps in an attempt to identify and understand pan-coronavirus molecular mechanisms. The use of a quantitative DIS allowed for the identification of virus-specific as well as shared interactions among distinct coronaviruses. We also systematically carried out subcellular localization analysis using tagged viral proteins and antibodies targeting specific SARS-CoV-2 proteins. Our results suggest that protein localization can often differ when comparing individually expressed viral proteins with the localization of the same protein in the context of infection. This can be because of factors such as mislocation driven by tagging, changes in localization due to interaction partners, or cellular compartments that are specific to the infection state. These differences are notable caveats of virus-host interaction studies performed with tagged, expressed proteins. However, previous studies and the work performed here show how these data can be powerful for the identification of host-targeted processes and relevant drug targets.

These data were integrated with genetic data where the interactions uncovered with SARS-CoV-2 were perturbed using RNA interference (RNAi) and CRISPR in different cellular systems and viral assays—an effort that functionally connected many host factors to infection. One of these, Tom70, which we have shown binds to ORF9b from both SARS-CoV-1 and SARS-CoV-2, is a mitochondrial outer membrane translocase that has been previously shown to be important for mounting an interferon response (42). Our functional data, however, show that Tom70 has at least some role in promoting infection rather than inhibiting it. Using cryo-EM, we obtained a 3-Å structure of a region of ORF9b binding to the active site of Tom70. Notably, we found that ORF9b is in a markedly different conformation than previously visualized. This suggests the possibility that ORF9b may partition between two distinct structural states, with each having a different function and possibly explaining its apparent pleiotropy. The exact details of functional significance and regulation of the ORF9b-Tom70 interaction will require further experimental elucidation. This interaction, however, which is conserved between SARS-CoV-1 and SARS-CoV-2, could have value as a pan-coronavirus therapeutic target.

Finally, we attempted to connect our in vitro molecular data to clinical information available for COVID-19 patients to understand the pathophysiology of COVID-19 and explore therapeutic avenues. To this end, using GWAS datasets of the COVID-HGI (30), we observed that increased predicted sIL17RA plasma levels were associated with lower risk of COVID-19. Notably, we find that IL17RA physically binds to SARS-CoV-2 ORF8, and genetic disruption results in decreased infection. These collective data suggest that future studies should be focused on this pathway as

both an indicator and therapeutic target for COVID-19. Furthermore, using medical billing data, we also observed trends in COVID-19 patients on specific drugs indicated by our molecular studies. For example, inpatients prescribed sigma-ligand typical antipsychotics appear to have better COVID-19 outcomes compared with users of atypical antipsychotics, which do not have anti–SARS-CoV-2 activity in vitro. However, we cannot be certain that the sigma receptor interaction is the mechanism underpinning this effect, as typical antipsychotics are known to bind to a multitude of cellular targets, and some atypical antipsychotics, which lack anti–SARS-CoV-2 activity, nonetheless have reported affinity for rodent sigma receptors (table S10). Replication in other patient cohorts and further work will be needed to see whether there is therapeutic value in these connections, but we have at least demonstrated a strategy wherein protein network analyses can be used to make testable predictions from real-world clinical information.

We have described an integrative and collaborative approach to study and understand pathogenic coronavirus infection, identifying conserved targeted mechanisms that are likely to be of high relevance for other viruses of this family, some of which have yet to infect humans. We used proteomics, cell biology, virology, genetics, structural biology, biochemistry, and clinical and genomic information in an attempt to provide a holistic view of SARS-CoV-2 and other coronaviruses' interactions with infected host cells. We propose that such an integrative and collaborative approach could and should be used to study other infectious agents as well as other disease areas.

Materials and methods

Cells

HEK293T/17 (HEK293T) cells were procured from the University of California, San Francisco (UCSF) Cell Culture Facility, and are available through UCSF's Cell and Genome Engineering Core (https://cgec.ucsf.edu/cell-culture-and-banking-services). HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM) (Corning) supplemented with 10% fetal bovine serum (FBS) (Gibco, Life Technologies) and 1% penicillin-streptomycin (Corning) and maintained at 37°C in a humidified atmosphere of 5% CO2. Short tandem repeat (STR) analysis by the Berkeley Cell Culture Facility on 8 August 2017 authenticates these as HEK293T cells with 94% probability.

HeLaM cells (RRID: CVCL_R965) were originally obtained from the laboratory of M. S. Robinson (CIMR, University of Cambridge, UK) and have been routinely tested for mycoplasma contamination. HeLaM cells were grown in DMEM supplemented with 10% FBS, 100 U/ml penicillin, 100 µg/ml streptomycin, and 2 mM glutamine at 37°C in a 5% CO2 humidified incubator.

A549 cells stably expressing ACE2 (A549-ACE2) were a gift from O. Schwartz. A549-ACE2 cells were cultured in DMEM supplemented with 10% FBS, blasticidin (20 µg/ml) (Sigma) and maintained at 37°C with 5% CO2. STR analysis by the Berkeley Cell Culture Facility on 17 July 2020 authenticates these as A549 cells with 100% probability.

Caco-2 cells (ATTC, HTB-37, RRID:CVCL_0025) were cultured in DMEM with GlutaMAX and pyruvate (Gibco, 10569010) and supplemented with 20% FBS (Gibco, 26140079). For Caco-2 cells utilized in Cas9-RNP knockouts, STR analysis by the Berkeley Cell Culture Facility on 23 April 2020 authenticates these as Caco-2 cells with 100% probability.

Vero E6 cells were purchased from the American Type Culture Collection (ATCC) and thus authenticated [VERO C1008 (Vero 76, clone E6, Vero E6)] (ATCC, CRL-1586). Vero E6 cells tested negative for mycoplasma contamination. Vero E6 cells were cultured in DMEM (Corning) supplemented with 10% FBS (Gibco, Life Technologies) and 1% penicillin-streptomycin (Corning) and maintained at 37°C in a humidified atmosphere of 5% CO2.

Microbes

LOBSTER E. coli Expression Strain: LOBSTR-[BL21(DE3)] Kerafast no. EC1002.

Antibodies

Commercially available primary antibodies used in this study

Rabbit anti-beta-actin (Cell Signaling Technology #4967, RRID:AB_330288); mouse anti-beta tubulin (Sigma-Aldrich #T8328, RRID:AB_1844090); rabbit anti-BiP (Cell Signaling Technology #3177S, RRID:AB_2119845); mouse anti-EEA1 (BD Biosciences #610457, RRID:AB_397830, used at 1:200); mouse anti-ERGIC53 (Enzo Life Sciences #ALX-804-602-C100, RRID:AB_2051363, used at 1:200); anti-GM130; rabbit anti-GRP78 BiP (Abcam #Ab21685, RRID:AB_2119834); rabbit anti–SARS-CoV–nucleocapsid protein (NP) (Rockland #200-401-A50, RRID:AB_828403); rabbit anti-PDI (protein disulfide isomerase) (Cell Signaling Technology #3501, RRID:AB_2156433); mouse anti-Strep tag (QIAGEN #34850, RRID:AB_2810987, used at 1:5000); mouse anti-strepMAB (IBA Lifesciences #2-1507-001, used at 1:1000); rabbit anti–Strep-tag II (Abcam #ab232586); rabbit anti-Tom20 (Proteintech #11802-1-AP, RRID:AB_2207530, used at 1:1000); rabbit anti-Tom20 (Cell Signaling Technology #42406, RRID:AB_2687663); mouse anti-Tom22 (Santa Cruz Biotechnology #sc-101286, RRID:AB_1130526); rabbit anti-Tom40 (Santa Cruz Biotechnology #sc-11414, RRID:AB_793274); mouse anti-Tom70 (Santa Cruz #sc-390545, RRID:AB_2714192, used at 1:500); Rabbit anti-STX5 (Synaptic Systems 110 053, used at 1:500); and ActinStaining Kit 647-Phalloidin (Hypernol #8817-01, used at 1:400).

Commercially available secondary antibodies used in this study

Alexa Fluor 488 chicken anti-mouse immunoglobulin G (IgG) (Invitrogen #A21200, RRID_AB_2535786, used at 1:400); Alexa Fluor 488 chicken anti-rabbit IgG (Invitrogen #A21441, RRID_AB_10563745, used at 1:400); Alexa Fluor 568 donkey anti-sheep IgG (Invitrogen #A21099, RRID_AB_10055702, used at 1:400); Alexa Fluor Plus 488 goat anti-rabbit (ThermoFisher A32731, used at 1:500); Alexa Fluor Plus 594 goat anti-mouse (ThermoFisher A32742, used at 1:500); and goat anti-mouse IgG-HRP (horseradish peroxidase) (BioRad #170-6516, RRID:AB_11125547, used at 1:20000).

Noncommercial antisera

Rabbit anti–SARS-CoV-2–NP antiserum was produced by the Garcia-Sastre laboratory and used at 1:10000. For information on polyclonal sheep antibodies targeting SARS-CoV-2 proteins, see below, table S3, and https://mrcppu-covid.bio/.

Coronavirus annotation and plasmid cloning

SARS-CoV-1 isolate Tor2 (NC_004718) and MERS-CoV (NC_019843) were downloaded from GenBank and utilized to design 2xStrep-tagged expression constructs of ORFs and proteolytically mature Nsps derived from ORF1ab (with N-terminal methionines and stop codons added as necessary). Protein termini were analyzed for predicted acylation motifs, signal peptides, and transmembrane regions, and either the N or C terminus was chosen for tagging as appropriate. Finally, reading frames were codon optimized and cloned into pLVX-EF1alpha-IRES-Puro (Takara/Clontech) including a 5′ Kozak motif.

Immunofluorescence microscopy of viral protein constructs

Approximately 60,000 HeLaM cells were seeded onto glass coverslips in a 12-well dish and grown overnight. The cells were transfected using 0.5 µg of plasmid DNA and either polyethylenimine (Polysciences) or Fugene HD (Promega; 1 part DNA to 3 parts transfection reagent) and grown for a further 16 hours.

Transfected cells were fixed with 4% paraformaldehyde (Polysciences) in phosphate-buffered saline (PBS) at room temperature for 15 min. The fixative was removed and quenched using 0.1 M glycine in PBS. The cells were permeabilized using 0.1% saponin in PBS containing 10% FBS. The cells were stained with the indicated primary and secondary antibodies for 1 hour at room temperature. The coverslips were mounted onto microscope slides using ProLong Gold antifade reagent (ThermoFisher) and imaged using a UplanApo 60x oil (NA 1.4) immersion objective on a Olympus BX61 motorized wide-field epifluorescence microscope. Images were captured using a Hamamatsu Orca monochrome camera and processed using ImageJ.

To gain insight into the intracellular distribution of each Strep-tagged construct, ~100 cells per transfection were manually scored. Each construct was assigned an intracellular distribution in relation to the plasma membrane, ER, Golgi, cytoplasm, and mitochondria (scored out of 7). Many of the constructs had several localizations so this was also reflected in the scoring. The scoring also took into account the impact of expression level on the localization of the constructs.

Meta-analysis of immunofluorescence data

We first sorted the data concerning viral protein location for all Strep-tagged viral proteins expressed individually in three heatmaps (one per virus) using a custom R script (“pheatmap” package). The information concerning protein localization during SARS-CoV-2 infection was added as a square border color code in the first heatmap, to compare the two different localization

patterns. To compare the predicted versus the experimentally determined locations, for each protein we took the top scoring sequence-based localization prediction from DeepLoc (43) if the score was >1. When more than one localization can be assigned to the same protein, we took as many top scoring ones as experimentally assigned localizations we had for the same protein. Finally, for each cell compartment, we count the number of experimentally assigned viral proteins and the subset of them predicted to that same compartment as correct predictions. To compare changes in protein interactions with changes in protein localization (Strep-tagged experiment versus sequence-based prediction), we calculated the Jaccard index of prey overlap for each viral protein (SARS-CoV-2 versus SARS-CoV-1 and SARS-CoV-2 versus MERS-CoV) and plotted them together, for proteins with the same localization and for proteins with different localization.

Generation of polyclonal sheep antibodies targeting SARS-CoV-2 proteins

Sheep were immunized with individual N-terminal glutathione S-transferase (GST)–tagged SARS-CoV-2 recombinant proteins or N-terminal maltose binding protein (MBP)–tagged proteins (for SARS-CoV-2 S, S-RBD, and ORF7a), followed by up to five booster injections 4 weeks apart from each other. Sheep were subsequently bled and IgGs were affinity purified using the specific recombinant N-terminal MBP-tagged viral proteins. Each antiserum specifically recognized the appropriate native viral protein. Characterization of each antibody can be found at https://mrcppu-covid.bio/. All antibodies generated can be requested at https://mrcppu-covid.bio/. Also see table S3.

Immunofluorescence microscopy of infected Caco-2 cells

For infection experiments in human colon epithelial Caco-2 cells (ATCC, HTB-37), SARS-CoV-2 isolate Muc-IMB-1 (provided by the Bundeswehr Institute of Microbiology, Munich, Germany) was used. SARS-CoV-2 was propagated in Vero E6 cells in DMEM supplemented with 2% FBS. All work involving live SARS-CoV-2 was performed in the BSL3 facility of the Institute of Virology, University Hospital Freiburg, and was approved according to the German Act of Genetic Engineering by the local authority (Regierungspraesidium Tuebingen, permit UNI.FRK.05.16/05).

Caco-2 human colon epithelial cells seeded on glass coverslips were infected with SARS-CoV-2 {Strain Muc-IMB-1/2020, second passage on Vero E6 cells [2 × 106 plaque-forming units (PFU)/ml]} at a multiplicity of infection (MOI) of 0.1. At 24 hours postinfection, cells were washed with PBS and fixed in 4% paraformaldehyde in PBS for 20 min at room temperature, followed by 5 min of quenching in 0.1 M glycine in PBS at room temperature. Cells were permeabilized and blocked in 0.1% saponin in PBS supplemented with 10% FBS for 45 min at room temperature

and incubated with primary antibodies for 1 hour at room temperature. After washing 15 min with blocking solution, AF568-labeled donkey–anti-sheep (Invitrogen, #A21099; 1:400) secondary antibody as well as AF4647-labeled Phalloidin (Hypermol, #8817-01; 1:400) were applied for 1 hour at room temperature. Subsequent washing was followed by embedding in Diamond Antifade Mountant with 4′,6-diamidino-2-phenylindole (DAPI) (ThermoFisher, #P36971). Fluorescence images were generated using a LSM800 confocal laser-scanning microscope (Zeiss) equipped with a 63X, 1.4 NA oil objective and Airyscan detector and the Zen blue software (Zeiss) and processed with Zen blue software and ImageJ/Fiji.

Transfection and cell harvest for immunoprecipitation experiments

For each affinity purification [SARS-CoV-1 baits, MERS-CoV baits, green fluorescent protein (GFP)–2xStrep, or empty vector controls], 10 million HEK293T cells were transfected with up to 15 μg of individual expression constructs using PolyJet transfection reagent (SignaGen Laboratories) at a 1:3 μg:μl ratio of plasmid to transfection reagent on the basis of the manufacturer's protocol. After >38 hours, cells were dissociated at room temperature using 10 ml PBS without calcium and magnesium (D-PBS) with 10 mM ethylenediaminetetraacetic acid (EDTA) for at least 5 min, pelleted by centrifugation at 200 × g, at 4°C for 5 min, washed with 10 ml D-PBS, pelleted once more, and frozen on dry ice before storage at −80°C for later immunoprecipitation analysis. For each bait, three independent biological replicates were prepared.

Whole-cell lysates were resolved on 4 to 20% Criterion SDS–polyacrylamide gel electrophoresis (SDS-PAGE) gels (Bio-Rad Laboratories) to assess Strep-tagged protein expression by immunoblotting using mouse anti-Strep tag antibody 34850 (QIAGEN) and anti-mouse HRP secondary antibody (BioRad).

Anti–Strep tag affinity purification

Frozen cell pellets were thawed on ice for 15 to 20 min and suspended in 1 ml lysis buffer [immunoprecipitation (IP) buffer (50 mM tris-HCl, pH 7.4 at 4°C; 150 mM NaCl, 1 mM EDTA) supplemented with 0.5% Nonidet P 40 Substitute (NP-40; Fluka Analytical) and cOmplete mini EDTA-free protease and PhosSTOP phosphatase inhibitor cocktails (Roche)]. Samples were then freeze-fractured by refreezing on dry ice for 10 to 20 min, then rethawed and incubated on a tube rotator for 30 min at 4°C. Debris was pelleted by centrifugation at 13,000 × g, at 4°C for 15 min. Up to 56 samples were arrayed into a 96-well Deepwell plate for affinity purification on the KingFisher Flex Purification System (Thermo Scientific) as follows: MagStrep "type3" beads (30

µl; IBA Lifesciences) were equilibrated twice with 1 ml wash buffer (IP buffer supplemented with 0.05% NP-40) and incubated with 0.95 ml lysate for 2 hours. Beads were washed three times with 1 ml wash buffer and then once with 1 ml IP buffer. Beads were released into 75 µl denaturation-reduction buffer [2 M urea, 50 mM Tris-HCl pH 8.0, 1 mM dithiothreitol (DTT)] in advance of on-bead digestion. All automated protocol steps were performed at 4°C using the slow mix speed and the following mix times: 30 s for equilibration and wash steps, 2 hours for binding, and 1 min for final bead release. Three 10-s bead collection times were used between all steps.

On-bead digestion for affinity purification

Bead-bound proteins were denatured and reduced at 37°C for 30 min, alkylated in the dark with 3 mM iodoacetamide for 45 min at room temperature, and quenched with 3 mM DTT for 10 min. To offset evaporation, 22.5 µl 50 mM Tris-HCl, pH 8.0 were added before trypsin digestion. Proteins were then incubated at 37°C, initially for 4 hours with 1.5 µl trypsin (0.5 µg/µl; Promega) and then another 1 to 2 hours with 0.5 µl additional trypsin. All steps were performed with constant shaking at 1100 rpm on a ThermoMixer C incubator. Resulting peptides were combined with 50 µl 50 mM Tris-HCl, pH 8.0 used to rinse beads and acidified with trifluoroacetic acid (0.5% final, pH < 2.0). Acidified peptides were desalted for MS analysis using a BioPureSPE Mini 96-Well Plate (20 mg PROTO 300 C18; The Nest Group, Inc.) according to standard protocols.

MS operation and peptide search

Samples were resuspended in 4% formic acid, 2% acetonitrile solution, and separated by a reversed-phase gradient over a nanoflow C18 column (Dr. Maisch). HPLC buffer A was composed of 0.1% formic acid, and HPLC buffer B was composed of 80% acetonitrile in 0.1% formic acid. Peptides were eluted by a linear gradient from 7 to 36% B over the course of 52 min, after which the column was washed with 95% B and re-equilibrated at 2% B. Each sample was directly injected by means of an Easy-nLC 1200 (Thermo Fisher Scientific) into a Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific) and analyzed with a 75-min acquisition, with all MS1 and MS2 spectra collected in the orbitrap; data were acquired using the Thermo software Xcalibur (4.2.47) and Tune (2.11 QF1 Build 3006). For all acquisitions, QCloud was used to control instrument longitudinal performance during the project (44). All proteomic data were searched against the human proteome (uniprot reviewed sequences downloaded 28 February 2020), enhanced green fluorescent protein (EGFP) sequence, and the SARS-CoV or MERS protein sequences using the default settings for MaxQuant (version 1.6.12.0) (45). Detected peptides and proteins were filtered to 1% false discovery rate in MaxQuant. All MS raw data and

search results files have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset (PXD identifier PXD021588).

High-confidence protein interaction scoring

Identified proteins were then subjected to PPI scoring with both SAINTexpress (version 3.6.3) and MiST (https://github.com/kroganlab/mist) (6, 7). We applied a two-step filtering strategy to determine the final list of reported interactors, which relied on two different scoring stringency cut-offs. In the first step, we chose all protein interactions that had a MiST score ≥0.7, a SAINTexpress Bayesian false-discovery rate (BFDR) ≤0.05, and an average spectral count ≥2. For all proteins that fulfilled these criteria, we extracted information about the stable protein complexes that they participated in from the CORUM (46) database of known protein complexes. In the second step, we then relaxed the stringency and recovered additional interactors that (i) formed complexes with interactors determined in filtering step 1 and (ii) fulfilled the following criteria: MiST score ≥0.6, SAINTexpress BFDR ≤0.05, and average spectral counts ≥2. Proteins that fulfilled filtering criteria in either step 1 or step 2 were considered to be high-confidence protein-protein interactions (HC-PPIs).

Using this filtering criteria, nearly all of our baits recovered a number of HC-PPIs in close alignment with previous datasets reporting an average of around six PPIs per bait (47). However, for a subset of baits, we observed a much higher number of PPIs that passed these filtering criteria. For these baits, the MiST scoring was instead performed using a larger in-house database of 87 baits that were prepared and processed in an analogous manner to this SARS-CoV-2 dataset. This was done to provide a more comprehensive collection of baits for comparison, to minimize the classification of nonspecifically binding background proteins as HC-PPIs. This was performed for SARS-CoV-1 baits (M, Nsp12, Nsp13, Nsp8, and ORF7b), MERS-CoV baits (Nsp13, Nsp2, and ORF4a), and SARS-CoV-2 Nsp16. SARS-CoV-2 Nsp16 MiST was scored using the in-house database as well as all previous SARS-CoV-2 data (5).

Hierarchical clustering of virus-human protein interactions

Hierarchical clustering was performed on interactions for (i) viral bait proteins shared across all three viruses and (ii) passed the high-confidence scoring criteria (MiST score ≥0.6, SAINTexpress BFDR ≤0.05, and average spectral counts ≥2) in at least one virus. We clustered using a new interaction score (K), which we defined as the average between the MiST and SAINT score for each virus-human interaction. This was done to provide a single score that captured the benefits from each scoring method. Clustering was performed using the ComplexHeatmap package in R,

using the “average” clustering method and “euclidean” distance metric. K-means clustering (k = 7) was applied to capture all possible combinations of interaction patterns between viruses.

GO enrichment analysis on clusters

Sets of genes found in seven clusters were tested for enrichment of GO terms, which was performed using the enricher function of clusterProfiler package in R (48). The GO terms were obtained from the C5 collection of Molecular Signature Database (MSigDBv7.1) and include biological process, cellular component, and molecular function ontologies. Significant GO terms were identified (adjusted $P < 0.05$) and further refined to select nonredundant terms. To select nonredundant gene sets, we first constructed a GO term tree based on distances (1 − Jaccard similarity coefficients of shared genes) between the significant terms. The GO term tree was cut at a specific level ($h = 0.99$) to identify clusters of nonredundant gene sets. For results with multiple significant terms belonging to the same cluster, we selected the term with the lowest adjusted P value.

Sequence similarity analysis

Protein sequence similarity was assessed by comparing the protein sequences from SARS-CoV-1 and MERS-CoV to SARS-CoV-2 for orthologous viral bait proteins. The corresponding PPI similarity was represented by a Jaccard index, using the high-confidence interactomes for each virus.

GO enrichment and PPI similarity analysis

The high-confidence interactors of the three viruses were tested for enrichment of GO terms as described above. We then identified GO terms that are significantly enriched (adjusted P value < 0.05) in all three viruses. For each enriched term, we generated the list of its associated genes and computed the Jaccard index of pairwise comparisons of the three viruses.

Orthologous versus nonorthologous interactions analysis

For a given pair of viruses, we identified all pairs of baits that share interactors and categorized these into orthologous and nonorthologous groups on the basis of whether the two baits were orthologs or not. We then summed up the total number of shared interactors in each group to calculate the corresponding fractions. This was performed for all pairwise combinations of the three viruses.

Structural modeling and comparison of MERS-CoV ORF4a and SARS-CoV-2 Nsp8

To obtain a sensitive sequence comparison between MERS-CoV ORF4a and SARS-CoV-2 Nsp8, we took into consideration their homologs. We first searched for homologs of these proteins in the UniRef30 database using hhblits (1 iteration, E-value cutoff 1 × 10−3) (49). Subsequently, the resulting alignments were filtered to include only sequences with at least 80% coverage to the corresponding query sequence, and hidden Markov models (HMMs) were created using hhmake. Finally, the HMMs of ORF4a and Nsp8 homologs were locally aligned using hhalign. The structure of ORF4a was predicted de novo using trRosetta (50). To provide greater coverage than that provided by experimental structures, SARS-CoV-2 Nsp8 was modeled using the structure of its SARS-CoV homolog as template (PDB ID: 2AHM) (51) using SWISS-MODEL (52). To search for local structural similarities between ORF4a and Nsp8, we used Geometricus, a structure embedding tool based on three-dimensional (3D) rotation invariant moments (53). This generates so-called shape-mers, analogous to sequence k-mers. The structures were fragmented into overlapping k-mers on the basis of the sequence (k = 20) and into overlapping spheres surrounding each residue (radius = 15 Å). To ensure that the similarities found between these distinct structures were significant, we used a high resolution of 7 to define the shape-mers. This resulted in the identification of four different shape-mers common to ORF4a and Nsp8. We aligned the entire ORF4a structure with residues 96 to 191 of the Nsp8 structure (i.e., after removal of the long N-terminal helix) using the Caretta structural alignment algorithm (54), using 3D rotation invariant moments (53) for initial superposition. We optimized parameters to maximize the Caretta score. The resulting alignment used k = 30, radius = 16 Å, gap open penalty = 0.05, gap extend penalty = 0.005, and had a root mean square deviation (RMSD) of 7.6 Å across 66 aligning residues.

DIS analysis

We computed a DIS for interactions that (i) originated from viral bait proteins shared across all three viruses and (ii) passed the high-confidence scoring criteria (see above) in at least one virus. We defined the DIS to be the difference between the interaction scores (K) from each virus. A DIS near 0 indicates that the interaction is confidently shared between the two viruses being compared, whereas a DIS near −1 or +1 indicates that the host-protein interaction is specific for one virus or the other. We computed a fourth DIS (SARS-MERS) by averaging K from SARS-CoV-1 and SARS-CoV-2 before calculating the difference with MERS-CoV. Here, a DIS near +1 indicates SARS-specific interactions (shared between SARS-CoV-1 and SARS-CoV-2 but absent in MERS-CoV), a DIS near −1 indicates MERS-specific interactions (present in MERS-CoV and

absent or lowly confident in both SARS-CoVs), and a DIS near 0 indicates interactions shared between all three viruses.

For each pairwise virus comparison, as well as the SARS-MERS comparison, the DIS was defined on the basis of cluster membership of interactions (Fig. 3A). For the SARS2-SARS1 comparison, interactions from every cluster except 5 were used, as those interactions are considered absent from both SARS-CoV-2 and SARS-CoV-1. For the SARS2-MERS comparison, interactions from all clusters except 3 were used. For the SARS1-MERS comparison, interactions from all clusters except 6 were used. For the SARS-MERS comparison, only interactions from clusters 2, 4, and 5 were used.

Network generation and visualization

PPI networks were generated in Cytoscape (55) and subsequently annotated using Adobe Illustrator. Host-host physical interactions, protein complex definitions, and biological process groupings were derived from CORUM (46), GO (biological process), and manually curated from literature sources. All networks were deposited in NDEx (56).

siRNA library and transfection in A549-ACE2 cells

An OnTargetPlus siRNA SMARTpool library (Horizon Discovery) was purchased targeting 331 of the 332 human proteins previously identified to bind SARS-CoV-2 (5) (PDE4DIP was not available for purchase and was excluded from the assay). This library was arrayed in a 96-well format, with each plate also including two nontargeting siRNAs and one siRNA pool targeting ACE2 (table S12). The siRNA library was transfected into A549 cells stably expressing ACE2 (A549-ACE2, provided by O. Schwartz), using Lipofectamine RNAiMAX reagent (Thermo Fisher). Briefly, 6 pmol of each siRNA pool were mixed with 0.25 µl RNAiMAX transfection reagent and OptiMEM (Thermo Fisher) in a total volume of 20 µl. After a 5 min incubation period, the transfection mix was added to cells seeded in a 96-well format. Twenty-four hours after transfection, the cells were subjected to SARS-CoV-2 infection, as described in the section Viral infection and quantification assay in A549-ACE2 cells, or incubated for 72 hours to assess cell viability using the CellTiter-Glo luminescent viability assay according to the manufacturer's protocol (Promega). Luminescence was measured in a Tecan Infinity 2000 plate reader, and percentage viability calculated relative to untreated cells (100% viability) and cells lysed with 20% ethanol or 4% formalin (0% viability), included in each experiment.

Viral infection and quantification assay in A549-ACE2 cells

Cells seeded in a 96-well format were inoculated with a SARS-CoV-2 stock (BetaCoV/France/IDF0372/2020 strain, generated and propagated once in Vero E6 cells and a gift from the National Reference Centre for Respiratory Viruses at Institut Pasteur, Paris, originally supplied through the European Virus Archive goes Global platform) at a MOI of 0.1 PFU per cell. After a 1-hour incubation period at 37°C, the virus inoculum was removed, and replaced by DMEM containing 2% FBS (Gibco, Thermo Fisher). Seventy-two hours postinfection, the cell culture supernatant was collected, heat inactivated at 95°C for 5 min, and used for RT-qPCR analysis to quantify viral genomes present in the supernatant. Briefly, SARS-CoV-2–specific primers targeting the N gene region: 5′-TAATCAGACAAGGAACTGATTA-3′ (forward) and 5′-CGAAGGTGTGACTTCCATG-3′ (reverse) (57) were used with the Luna Universal One-Step RT-qPCR Kit (New England Biolabs) in an Applied Biosystems QuantStudio 6 thermocycler, with the following cycling conditions: 55°C for 10 min, 95°C for 1 min, and 40 cycles of 95°C for 10 s, followed by 60°C for 1 min. The number of viral genomes is expressed as PFU equivalents per milliliter, and was calculated by performing a standard curve with RNA derived from a viral stock with a known viral titer.

Knockdown validation with RT-qPCR in A549-ACE2 cells

Gene-specific qPCR primers targeting all genes represented in the OnTargetPlus library were purchased and arrayed in a 96-well format identical to that of the siRNA library (IDT; table S13). A549-ACE2 cells treated with siRNA were lysed using the Luna Cell Ready Lysis Module (New England Biolabs) following the manufacturer's protocol. The lysate was used directly for gene quantification by RT-qPCR with the Luna Universal One-Step RT-qPCR Kit (New England Biolabs), using the gene-specific PCR primers and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as a housekeeping gene. The following cycling conditions were used in an Applied Biosystems QuantStudio 6 thermocycler: 55°C for 10 min, 95°C for 1 min, and 40 cycles of 95°C for 10 s, followed by 60°C for 1 min. The fold change in gene expression for each gene was derived using the $2^{-\Delta\Delta CT}$, 2 (Delta Delta CT) method (58), normalized to the constitutively expressed housekeeping gene GAPDH. Relative changes were generated comparing the control siRNA knockdown transfected cells to the cells transfected with each siRNA.

Single guide RNA selection for Cas9 knockout screen

Single guide RNAs (sgRNAs) were designed according to Synthego's multiguide gene knockout (59). Briefly, two or three sgRNAs are bioinformatically designed to work in a cooperative manner

to generate small, knockout-causing, fragment deletions in early exons (fig. S18). These fragment deletions are larger than standard indels generated from single guides. The genomic repair patterns from a multiguide approach are highly predictable on the basis of the guide spacing and design constraints to limit off-targets, resulting in a higher probability protein knockout phenotype (table S14).

sgRNA synthesis for Cas9 knockout screen

RNA oligonucleotides were chemically synthesized on Synthego solid-phase synthesis platform, using CPG solid support containing a universal linker. 5-benzylthio-1H-tetrazole (BTT, 0.25 M solution in acetonitrile) was used for coupling, [3-((dimethylamino-methylidene)amino)-3H-1,2,4-dithiazole-3-thione (DDTT, 0.1 M solution in pyridine)] was used for thiolation, dichloroacetic acid (DCA, 3% solution in toluene) was used for detritylation. Modified sgRNA were chemically synthesized to contain 2′-O-methyl analogs and 3′ phosphorothioate nucleotide interlinkages in the terminal three nucleotides at both 5′ and 3′ ends of the RNA molecule. After synthesis, oligonucleotides were subject to a series of deprotection steps, followed by purification by solid-phase extraction (SPE). Purified oligonucleotides were analyzed by electrospray ionization mass spectrometry (ESI-MS).

Arrayed knockout generation with Cas9-RNPs

For Caco-2 transfection, 10 pmol Streptococcus Pyogenes NLS-Sp.Cas9-NLS (SpCas9) nuclease (Aldevron; 9212) was combined with 30 pmol total synthetic sgRNA (10 pmol each sgRNA, Synthego) to form ribonucleoproteins (RNPs) in 20 µl total volume with SF Buffer (Lonza V5SC-2002) and allowed to complex at room temperature for 10 min.

All cells were dissociated into single cells using TrypLE Express (Gibco), resuspended in culture media and counted. 100,000 cells per nucleofection reaction were pelleted by centrifugation at 200 × g for 5 min. After centrifugation, cells were resuspended in transfection buffer according to cell type and diluted to 2 × 104 cells/µl. Five µl of cell solution was added to preformed RNP solution and gently mixed. Nucleofections were performed on a Lonza HT 384-well nucleofector system (Lonza, #AAU-1001) using program CM-150 for Caco-2. Immediately after nucleofection, each reaction was transferred to a tissue-culture treated 96-well plate containing 100 µl of normal culture media and seeded at a density of 50,000 cells per well. Transfected cells were incubated following standard protocols.

Quantification of arrayed knockout efficiency

Two days after nucleofection, genomic DNA was extracted from cells using DNA QuickExtract (Lucigen, #QE09050). Briefly, cells were lysed by removal of the spent media followed by addition of 40 µl of QuickExtract solution to each well. Once the QuickExtract DNA Extraction Solution was added, the cells were scraped off the plate into the buffer. After transfer to compatible plates, DNA extract was then incubated at 68°C for 15 min followed by 95°C for 10 min in a thermocycler before being stored for downstream analysis.

Amplicons for indel analysis were generated by PCR amplification with NEBNext polymerase (NEB, #M0541) or AmpliTaq Gold 360 polymerase (Thermo Fisher Scientific, #4398881) according to the manufacturer's protocol. The primers were designed to create amplicons between 400 and 800 base pairs (bp), with both primers at least 100 bp from any of the sgRNA target sites (table S15). PCR products were cleaned-up and analyzed by Sanger sequencing (Genewiz). Sanger data files and sgRNA target sequences were input into Inference of CRISPR Edits (ICE) analysis (ice.synthego.com) to determine editing efficiency and to quantify generated indels (60). Percentage of alleles edited is expressed as an ice-d score. This score is a measure of how discordant the sanger trace is before versus after the edit. It is a simple and robust estimate of editing efficiency in a pool, especially suited to highly disruptive editing techniques like multiguide.

Identification of essential genes for siRNA and Cas9 knockout screen

We used longitudinal imaging in A549 cells to assess cell viability (fig. S18). For benchmarking, relative cell viability was measured by CellTiter-Glo Luminescent Cell Viability Assay (Promega; G7571) as per manufacturer's instructions. Briefly, two passages postnucleofection A549 siRNA pools cultured in 96-well tissue-culture treated plates (Corning, #3595) were lysed in the CellTIter-Glo reagent, by removing spent media and adding 100 µl of the CellTiter-Glo reagent containing the CellTiter-Glo buffer and CellTiter-Glo Substrate. Cells were placed on an orbital shaker for 2 min on a SpectraMax iD5 (Molecular Devices) and then incubated in the dark at room temperature for 10 min. Completely lysed cells were pipette mixed and 25 µl were transferred to a 384-well assay plate (Corning, #3542). The luminescence was recorded on a SpectraMax iD5 (Molecular Devices) with an integration time of 0.25 s per well. Luminescence readings were all normalized to the without-sgRNA control condition.

To determine cell viability in Caco-2 knockouts we used longitudinal imaging (fig. S18). All gene knockout pools were maintained for a minimum of six passages to determine the effect of loss of

protein function on cell fitness before viral infection. Viability was determined through longitudinal imaging and automated image analysis using a Celigo Imaging Cytometer (Celigo). Each gene knockout pool was split in triplicate wells on separate plates. Every day, except the day of seeding, each well was scanned and analyzed using built-in Confluence imaging parameters using autoexposure and autofocus with an offset of −45 μm. Analysis was performed with standard settings except for an intensity threshold setting of 8. Confluency was averaged across three wells and plotted over time. Viability genes were determined as pools that, after six passages, remained <20% confluent 5 days after seeding. Genes deemed essential were excluded from the knockout screen.

Cells, virus, and infections for Caco-2 Cas9 knockout screen

Wild-type and CRISPR-edited Caco-2 cells were grown at 37°C, 5% CO2 in DMEM, 10% FBS. SARS-CoV-2 stocks were grown and titered on Vero E6 cells as described previously (61). Wild-type and CRISPR-edited Caco-2 cell lines were infected with SARS-CoV-2 at a MOI of 0.01 in DMEM supplemented with 2% FBS. Seventy-two hours postinfection, supernatants were harvested and stored at −80°C and the Caco-2 wild-type (WT) and CRISPR knockout (KO) cells were fixed with 10% neutral buffered formalin (NBF) for 1 hour at room temperature to enable further analysis.

Focus-forming assay for Caco-2 Cas9 knockout screen

Vero E6 cells were plated into 96-well plates at confluence (50,000 cells per well) in DMEM supplemented with 10% heat-inactivated FBS (Gibco). Before infection, supernatants from infected Caco-2 WT and CRISPR KO cells were thawed and serially diluted from 10−1 to 10−8. Growth media was removed from the Vero E6 cells and 40 μl of each virus dilution was plated. After 1 hour of adsorption at 37°C, 5% CO2, 40 μl of 2.4% microcrystalline cellulose (MCC) overlay supplemented with DMEM powdered media (Gibco) to a concentration of 1x was added to each well of the 96-well plate to achieve a final MCC overlay concentration of 1.2%. Plates were then incubated at 37°C, 5% CO2 for 24 hours. The MCC overlay was gently removed and cells were fixed with 10% NBF for 1 hour at room temperature. After removal of NBF, monolayers were washed with ultrapure water and ice-cold 100% methanol/0.3% H2O2 was added for 30 min to permeabilize the cells and quench endogenous peroxidase activity. Monolayers were then blocked for 1 hour in PBS with 5% nonfat dry milk (NFDM). After blocking, monolayers were incubated with SARS-CoV N primary antibody (Novus Biologicals; NB100-56576; 1:2000) for 1 hour at room temperature in PBS, 5% NFDM. Monolayers were washed with PBS and incubated

with an HRP-conjugated secondary antibody for 1 hour at room temperature in PBS with 5% NFDM. Secondary antibody was removed, monolayers were washed with PBS, and then developed using TrueBlue substrate (KPL) for 30 min. Plates were imaged on a Bio-Rad Chemidoc utilizing a phosphorscreen and foci were counted by eye to calculate focus-forming units per ml (FFU/ml) for each knockout. The original formalin-fixed Caco-2 WT and CRISPR KO cells were stained with DAPI (Thermo Scientific) and imaged on a Cytation 5-plate reader to determine cell viability. Wells containing no cells were excluded from further analyses.

Quantitative analysis and scoring of knockdown and knockout library screens

Virus readout by qPCR (A549-ACE2, expressed as plaque-forming units per milliliter) and focus-forming assay readouts (Caco-2, focus-forming units per milliliter) were processed using the RNAither package (www.bioconductor.org/packages/release/bioc/html/RNAither.html) in the statistical computing environment R. The two datasets were normalized separately, using the following method. The readouts were first log transformed (natural logarithm), and robust z-scores [using median and MAD (median absolute deviation) instead of mean and standard deviation] were then calculated for each 96-well plate separately. z-scores of multiple replicates of the same perturbation were averaged into a final z-score for presentation in Fig. 5. No filtering was done on the basis of differences in replicate z-scores, but all replicate scores are individually listed in tables S6 and S7. We suggest consulting the replicate z-scores for all genes and perturbations of interest. The A549-ACE2 siRNA screen includes three replicates (or more) of each perturbation, and the Caco-2 CRISPR screen includes two replicates (or more) of each perturbation. The results from the A549-ACE2 screen cover all 332 screened genes (331 SARS-CoV-2 interactors plus ACE2). The results from the Caco-2 screen cover 286 of the screened genes plus ACE2. The remaining Caco-2 genes were either deemed essential, failed editing, or failed in the focus-forming assay.

Antiviral drug and cytotoxicity assays (A549-ACE2 cells)

In total, 2500 A549-ACE2 cells were seeded into 96- or 384-well plates in DMEM (10% FBS) and incubated for 24 hours at 37°C, 5% CO2. Two hours before infection, the media was replaced with 120 µl (96-well format) or 50 µl (384-well format) of DMEM (2% FBS) containing the compound of interest at the indicated concentration. At the time of infection, the media was replaced with virus inoculum (MOI 0.1 PFU per cell) and incubated for 1 hour at 37°C, 5% CO2. After the adsorption period, the inoculum was removed, replaced with 120 µl (96-well format) or 50 µl (384-well format) of drug-containing media, and cells were incubated for an additional 72

hours at 37°C, 5% CO2. At this point, the cell culture supernatant was harvested, and viral load was assessed by RT-qPCR (as described in the section Viral infection and quantification assay in A549-ACE2 cells). Viability was assayed using the CellTiter-Glo assay following the manufacturer's protocol (Promega). Luminescence was measured in a Tecan Infinity 2000 plate reader, and percentage viability calculated relative to untreated cells (100% viability) and cells lysed with 20% ethanol or 4% formalin (0% viability), included in each experiment.

Antiviral drug and cytotoxicity assays (Vero E6 cells)

Viral growth and cytotoxicity assays in the presence of inhibitors were performed as previously described (5). In total, 2000 Vero E6 cells were seeded into 96-well plates in DMEM (10% FBS) and incubated for 24 hours at 37°C, 5% CO2. Two hours before infection, the medium was replaced with 100 µl of DMEM (2% FBS) containing the compound of interest at concentrations 50% greater than those indicated, including a DMSO control. SARS-CoV-2 virus (100 PFU; MOI 0.025) was added in 50 µl of DMEM (2% FBS), bringing the final compound concentration to those indicated. Plates were then incubated for 48 hours at 37°C. After infection, supernatants were removed, and cells were fixed with 4% formaldehyde for 24 hours before being removed from the BSL3 facility. The cells were then immunostained for the viral NP protein (rabbit antisera produced in the Garcia-Sastre laboratory; 1:10,000) with a DAPI counterstain. Infected cells (488 nm) and total cells (DAPI) were quantified using a Celigo (Nexcelcom) imaging cytometer. Infectivity is measured by the accumulation of viral NP protein in the nucleus of the cells (fluorescence accumulation). Percent infection was quantified as {[(number of infected cells / total cells) − background] × 100}, and the DMSO control was then set to 100% infection for analysis. The IC50 and IC90 for each experiment was determined using the Prism (GraphPad Software) software. Cytotoxicity measurements were performed using the MTT assay (Roche), according to the manufacturer's instructions. Cytotoxicity was performed in uninfected Vero E6 cells with same compound dilutions and concurrent with viral replication assay. All assays were performed in biologically independent triplicates. Sourcing information for all drugs tested may be found in table S10.

Coimmunoprecipitation assays for ORF9b and Tom70

HEK293T and A549 cells were transfected with the indicated mammalian expression plasmids using Lipofectamine 2000 (Invitrogen) and TransIT-X2 (Mirus Bio), respectively. Twenty-four hours after transfection, cells were harvested and lysed in NP-40 lysis buffer [0.5% Nonidet P 40 Substitute (NP-40; Fluka Analytical), 50 mM Tris-HCl, pH 7.4 at 4°C, 150 mM NaCl, 1 mM EDTA]

supplemented with cOmplete mini EDTA-free protease and PhosSTOP phosphatase inhibitor cocktails (Roche). Clarified cell lysates were incubated with Streptactin Sepharose beads (IBA) for 2 hours at 4°C, followed by five washes with NP-40 lysis buffer. Protein complexes were eluted in the SDS loading buffer and were analyzed by Western blotting with the indicated antibodies.

Quantification of Tom70 down-regulation in HeLaM cells overexpressing ORF9b

HeLaM cells were transiently transfected with plasmids encoding GFP-Strep, SARS-CoV-1 ORF9b-Strep, or SARS-CoV-2 ORF9b-Strep. The next day, the cells were fixed using 4% paraformaldehyde and immunostained with antibodies against Strep tag, and Tom20 or Tom70. Representative images for each construct were captured by acquiring a single optical section using a Nikon A1 confocal fitted with a CFI Plan Apochromat VC 60x oil objective (NA 1.4). For image quantification multiple fields of view were captured for each construct using a CFI Super Plan Fluor ELWD 40x objective (NA 0.6). The mean fluorescent intensity for Tom20 and Tom70 was measured by manually drawing a region of interest around each cell using ImageJ. Between 30 and 60 cells were quantified for each construct.

Quantification of Tom70 down-regulation in infected Caco-2 cells

Caco-2 cells were seeded on glass coverslips in triplicate and infected with SARS-CoV-2 at a MOI of 0.1 as described above. At 24 hours postinfection, cells were fixed with 4% paraformaldehyde and immunostained with antibodies against Tom70, Tom20, and ORF9b. For signal quantification images of noninfected and neighboring infected cells were acquired using a LSM800 confocal laser-scanning microscope (Zeiss) equipped with a 63X, 1.4 NA oil objective and the Zen blue software (Zeiss). The mean fluorescence intensity of each cell was measured by ImageJ software. Forty-three cells were quantified for each condition—infected or noninfected—from three independent experiments.

Coexpression and purification of ORF9b-Tom70 (residues 109 to end) complexes

SARS-CoV-2 ORF9b and Tom70 (residues 109 to end) were coexpressed using a pET29-b(+) vector backbone where ORF9b was tag-less and Tom70 had an N-terminal 10XHis-tag and SUMO-tag. LOBSTR E. coli cells transformed with the above construct were grown at 37°C until they reached an optical density at 600 nm (OD600) of 0.8, then expression was induced at 37°C with 1 mM IPTG for 4 hours. Frozen cell pellets were resuspended in 25 ml of lysis buffer (200 mM NaCl, 50 mM Tris-HCl pH 8.0, 10% v/v glycerol, 2 mM $MgCl_2$) per liter of cell culture, supplemented with cOmplete protease inhibitor tablets (Roche), 1 mM phenylmethylsulfonyl

fluoride (PMSF) (Sigma), 100 µg/ml lysozyme (Sigma), 5 µg/ml DNaseI (Sigma), and then homogenized with an immersion blender (Cuisinart). Cells were lysed by 3x passage through an Emulsiflex C3 cell disruptor (Avestin) at ~103,000 kPa, and the lysate clarified by ultracentrifugation at 100,000 × g for 30 min at 4°C. The supernatant was collected, supplemented with 20 mM imidazole, loaded into a gravity flow column containing Ni-NTA superflow resin (Qiagen), and rocked with the resin at 4°C for 1 hour. After allowing the column to drain, resin was rinsed twice with 5 column volumes (cv) of wash buffer [150 mM KCl, 30 mM Tris-HCl pH 8.0, 10% v/v glycerol, 20 mM imidazole, 0.5 mM tris(hydroxypropyl)phosphine (THP, VWR)] supplemented with 2 mM ATP (Sigma) and 4 mM MgCl2, then washed with 5 cv wash buffer with 40 mM imidazole. Resin was then rinsed with 5 cv Buffer A (50 mM KCl, 30 mM Tris-HCl pH 8.0, 5% glycerol, 0.5 mM THP) and protein was eluted with 2 × 2.5 cv Buffer A plus 300 mM imidazole. Elution fractions were combined, supplemented with Ulp1 protease, and rocked at 4°C for 2 hours. Ulp1-digested Ni-NTA eluate was diluted 1:1 with additional Buffer A, loaded into a 50 ml Superloop, and applied to a MonoQ 10/100 column on an Äkta pure system (GE Healthcare) using 100% Buffer A, 0% Buffer B (1000 mM KCl, 30 mM Tris-HCl pH 8.0, 5% glycerol, 0.5 mM THP). The MonoQ column was washed with 0 to 40% Buffer B gradient over 15 cv, peak fractions were analyzed by SDS-PAGE and the identity of tagless Tom70 (109 to end) and ORF9b proteins confirmed by intact protein MS (Xevo G2-XS Mass Spectrometer, Waters). Peak fractions eluting at ~15% B contained relatively pure Tom70 (109 to end) and ORF9b, and these were concentrated using 10 kDa Amicon centrifugal filter (Millipore) and further purified by size exclusion chromatography using a Superdex 200 increase 10/300 GL column (GE Healthcare) in buffer containing 150 mM KCl, 20 mM HEPES-NaOH pH 7.5, 0.5 mM THP. The sole size-exclusion peak contained both Tom70 (109 to end) and ORF9b, and the center fraction was used directly for cryo-EM grid preparation.

Expression and purification of SARS-CoV-2 ORF9b

ORF9b with N-terminal 10XHis-tag and SUMO-tag was expressed using a pET-29b(+) vector backbone. LOBSTR E. coli cells transformed with the above construct were grown at 37°C until they reached an optical density at 600 nm (OD600) of 0.8, then expression was induced at 37°C with 1 mM IPTG for 6 hours. Frozen cell pellets were lysed, homogenized, clarified, and subject to Ni affinity purification as described above for ORF9b-Tom70 complexes, with several small changes. Lysis buffers and Ni-NTA wash buffers contained 500 mM NaCl, and an additional wash step using 10 cv wash buffer plus 0.2% TWEEN20 plus 500 mM NaCl was carried out before the ATP wash. ORF9b was eluted from Ni-NTA resin in Buffer A (50 mM NaCl, 25 mM Tris pH 8.5,

5% glycerol, 0.5 mM THP) supplemented with 300 mM imidazole. This eluate was diluted 1:1 with additional Buffer A, loaded into a 50 ml Superloop, and applied to a MonoQ 10/100 column on an Äkta pure system (GE Healthcare) using 100% Buffer A, 0% Buffer B (1000 mM NaCl, 25mM Tris-HCl pH 8.5, 5% glycerol, 0.5 mM THP). The MonoQ column was washed with 0 to 40% Buffer B gradient over 15 cv, and relatively pure ORF9b eluted at 20 to 25% Buffer B, whereas ORF9b and contaminating proteins eluted at 30 to 35% buffer B. Fractions from these two peaks were combined and incubated with Ulp1 and HRV3C proteases at 4°C for 2 hours, supplemented with 10 mM imidazole, then thrice flowed back through 1 ml of Ni-NTA resin equilibrated with size-exclusion buffer (as above) plus10 mM imidazole. The reverse-Ni purified sample was concentrated using 10 kDa Amicon centrifugal filter and then further purified by size exclusion chromatography using a Superdex 200 increase 10/300 GL column.

Expression and purification of Tom70 (109-end)

Tom70 (109 to end) with N-terminal 10XHis-tag and SUMO-tag and C terminus Spy-tag, HRV-3C protease cleavage site, and eGFP-tag was expressed using a pET-21(+) vector backbone. LOBSTR E. coli cells transformed with the above construct were grown at 37°C until they reached an optical density at 600 nm (OD600) of 0.8, then expression was induced at 16°C with 0.5 mM IPTG overnight. The soluble domain of Tom70 [Tom70 (109-end)] was purified as described in (62) with some modifications. Frozen cell pellets of LOBSTR E. coli transformed with the above construct were resuspended in 50 ml lysis buffer (500 mM NaCl, 20 mM KH2PO4 pH 7.5) per liter cell culture, supplemented with 1 mM PMSF (Sigma) and 100 µg/ml, and homogenized. Cells were lysed by 3x passage through an Emulsiflex C3 cell disruptor (Avestin) at ~103,000 kPa, and the lysate clarified by ultracentrifugation at 100,000 × g for 30 min at 4°C. The supernatant was collected, supplemented with 20 mM imidazole, loaded into a gravity flow column containing Ni-NTA superflow resin (Qiagen), and rocked with the resin at 4°C for 1 hour. After allowing the column to drain, resin was rinsed twice with 5 column volumes (cv) of wash buffer (500 mM KCl, 20 mM KH2PO4 pH 8.0, 20 mM imidazole, 0.5 mM THP) supplemented with 2 mM ATP (Sigma) and 4 mM MgCl2, then washed with 5 cv wash buffer with 40 mM imidazole. Bound Tom70 (109 to end) was then cleaved from the resin by 2-hour incubation with Ulp1 protease in 4 cv elution buffer (150 mM KCl, 20 mM KH2PO4 pH 8.0, 5 mM imidazole, 0.5 mM THP). After cleavage with Ulp1, the flow through was collected along with a 2-cv rinse of the resin with additional elution buffer. These fractions were combined and HRV3C protease was added to remove the C-terminal EGFP tag (1:20 HRV3C to Tom70). After 2-hour HRV3C digestion at 4°C, the double-digested Tom70 (109 to end) was concentrated using a 30 kDa Amicon centrifugal filter (Millipore) and

further purified by size exclusion chromatography using a Superdex 200 increase 10/300 GL column (GE Healthcare) in buffer containing 150 mM KCl, 20 mM HEPES-NaOH pH 7.5, 0.5 mM THP.

Prediction of SARS-CoV-2 ORF9b internal mitochondrial targeting sequence

ORF9b was analyzed for the presence of an internal mitochondrial targeting sequence (i-MTS) as described in (63) using the TargetP-2.0 server (64). Sequences corresponding to ORF9b N-terminal truncations of 0 to 62 residues were submitted to the TargetP-2.0 server, and the probability of the peptides containing an MTS plotted against the numbers of residues truncated. A similar analysis using the MitoFates server (65) predicted that ORF9b residues 54 to 63 were the most likely to make up a presequence MTS on the basis of their propensity to form a positively charged amphipathic helix. Notably this analysis was consistent with the secondary structure prediction from JPRED (66).

Cryo-EM sample preparation and data collection

Three µl of ORF9b-Tom70 complex (12.5 µM) was added to a 400 mesh 1.2/1.3R Au Quantifoil grid previously glow discharged at 15 mA for 30 s. Blotting was performed with a blot force of 0 for 5 s at 4°C and 100% humidity in a FEI Vitrobot Mark IV (ThermoFisher) before plunge freezing into liquid ethane. A total of 1534 118-frame super-resolution movies were collected with a 3 by 3 image-shift collection strategy at a nominal magnification of 105,000x (physical pixel size: 0.834 Å per pixel) on a Titan Krios (ThermoFisher) equipped with a K3 camera and a Bioquantum energy filter (Gatan) set to a slit width of 20 eV. Collection dose rate was 8 electrons per pixel per second for a total dose of 66 electrons per square angstrom. Defocus range was −0.7 to −2.4 µm. Each collection was performed with semiautomated scripts in SerialEM (67).

Cryo-EM image processing and model building

We motion corrected 1534 movies using Motioncor2 (68), and we imported dose-weighted summed micrographs in cryosparc (version 2.15.0). Then, 1427 micrographs were curated on the basis of contrast transfer function (CTF) fit (better than 5 Å) from a patch CTF job. Template-based particle picking resulted in 2,805,121 particles, and 1,616,691 particles were selected after 2D classification. Five rounds of 3D classification using multiclass ab initio reconstruction and heterogeneous refinement yielded 178,373 particles. Homogeneous refinement of these final particles led to a 3.1-Å electron density map which was used for model building. The

reconstruction was filtered by the masked Fourier shell correlation (FSC) and sharpened with a b-factor of −145.

To build the model of Tom70 (109 to end), the crystal structure of Saccharomyces cerevisiae Tom71 (PDB ID: 3fp3; sequence identity 25.7%) was first fit into the cryo-EM density as a rigid body in UCSF ChimeraX and then relaxed into the final density using Rosetta FastRelax mover in torsion space. This model, along with a BLAST alignment of the two sequences (69), was used as a starting point for manual building using COOT (70). After initial building by hand, the regions with poor density fit or geometry were iteratively rebuilt using Rosetta (71). ORF9b was built de novo into the final density using COOT, informed and facilitated by the predictions of the TargetP-2.0, MitoFates, and JPRED servers. The ORF9b-Tom70 complex model was submitted to the Namdinator web server (72) and further refined in ISOLDE 1.0 (73) using the plugin for UCSF ChimeraX (74). Final model b-factors were estimated using Rosetta. The model was validated using phenix.validation_cryoem (75). The final model contains residues 109 to 272 and 298 to 600 of human Tom70 and 39 to 76 of SARS-CoV-2 ORF9b. Molecular interface between ORF9b and Tom70 was analyzed using the PISA web server (76). Figures were prepared using UCSF ChimeraX.

Computational human genetics analysis

To look for genetic variants associated with our list of proteins that had a meaningful impact on SARS-CoV-2 replication, we used the largest proteomic GWAS study to date (28). We identified IL17RA as one of the proteins assayed in Sun et al.'s proteomic GWAS and observed that it had multiple cis-acting protein quantitative trait loci (pQTLs) at a corrected P value of 1 × 10−5, where cis-acting is defined as within 1 Mb of the transcription start site of IL17RA.

We used the GSMR method (29) to perform Mendelian randomization (MR) using near-independent [linkage disequilibrium (LD) R2 = 0.05, where R2 is the coefficient of determination] cis-pQTLs for IL17RA. The advantage of the GSMR method over conventional MR methods is twofold. First, GSMR performs MR adjusting for any residual correlation between selected genetic variants by default. Second, GSMR has a built-in method called HEIDI (heterogeneity in dependent instruments)–outlier that performs heterogeneity tests in the near-independent genetic instruments and removes potentially pleiotropic instruments (i.e., where there is evidence of heterogeneity at $P < 0.01$). Details of the GSMR and HEIDI method have been published previously (29).

Summary statistics generated by COVID-HGI (round 3; www.covid19hg.org/results/) for COVID-19 versus population, hospitalized COVID-19 versus population and hospitalized COVID-19 versus nonhospitalized COVID-19 were used for IL17RA MR analysis. We used the 1000 genomes phase 3 European population genotype data to derive the LD correlation matrix for this analysis. The phenotype definitions as provided by COVID-HGI are as follows. COVID-19 versus population: Case, individuals with laboratory confirmation of SARS-CoV-2 infection, EHR/ICD coding/Physician-confirmed COVID-19, or self-reported COVID-19 positive; control, everybody that is not a case. Hospitalized COVID-19 versus population: case, hospitalized, laboratory confirmed SARS-CoV-2 infection or hospitalization due to COVID-19-related symptoms; control, everybody that is not a case, e.g., population. Hospitalized COVID-19 versus nonhospitalized COVID-19: case, hospitalized, laboratory confirmed SARS-CoV-2 infection or hospitalization due to COVID-19-related symptoms; control, laboratory confirmed SARS-CoV-2 infection and not hospitalized 21 days after the test.

Infections and treatments for IL-17A treatment studies

The WA-1 strain (BEI resources) of SARS-CoV-2 was used for all experiments. All live virus experiments were performed in a BSL3 laboratory. SARS-CoV-2 stocks were passaged in Vero E6 cells (ATCC) and titer was determined via plaque assay on Vero E6 cells as previously described (77). Briefly, virus was diluted 1:102 to 1:106 and incubated for 1 hour on Vero E6 cells before an overlay of Avicel and complete DMEM (Sigma Aldrich, SLM-241) was added. After incubation at 37°C for 72 hours, the overlay was removed and cells were fixed with 10% formalin, stained with crystal violet, and counted for plaque formation. SARS-CoV-2 infections of A549-ACE2 cells were done at a MOI of 0.05 for 24 hours. Inhibitors and cytokines were added concurrently with virus. All infections were done in technical triplicate. Cells were treated with the following compounds: Remdesivir (SELLECK CHEMICALS LLC, S8932) and IL-17A (Millipore-Sigma, SRP0675).

RNA extraction, RT, and RT-qPCR for IL-17A treatment studies

Total RNA from samples was extracted using the Direct-zol RNA kit (Zymogen, R2060) and quantified using the NanoDrop 2000c (ThermoFisher). cDNA was generated using 500 ng of RNA from infected A549-ACE2 cells with Superscript III reverse transcription (ThermoFisher, 18080-044) and oligo(dT)12-18 (ThermoFisher, 18418-012) and random hexamer primers (ThermoFisher, S0142). RT-qPCR reactions were performed on a CFX384 (BioRad) and delta cycle threshold (ΔCt) was determined relative to RPL13A levels. Viral detection levels and target

host genes in treated samples were normalized to water-treated controls. The SYBR green qPCR reactions contained 5 µl of 2x Maxima SYBR green/Rox qPCR Master Mix (ThermoFisher; K0221), 2 µl of diluted cDNA, and 1 nmol of both forward and reverse primers, in a total volume of 10 µl. The reactions were run as follows: 50°C for 2 min and 95°C for 10 min, followed by 40 cycles of 95°C for 5 s and 62°C for 30 s. Primer efficiencies were ~100%. Dissociation curve analysis after the end of the PCR confirmed the presence of a single and specific product. RT-qPCR primers were used against the SARS-CoV-2 E gene (PF_042_nCoV_E_F: ACAGGTACGTTAATAGTTAATAGCGT; PF_042_nCoV_E_R: ATATTGCAGCAGTACGCACACA), the CXCL8 gene (CXCL8 For: ACTGAGAGTGATTGAGAGTGGAC; CXCL8 Rev: AACCCTCTGCACCCAGTTTTC), and the RPL13A gene (RPL13A For: CCTGGAGGAGAAGAGGAAAGAGA; RPL13A Rev: TTGAGGACCTCTGTGTATTTGTCAA).

Transfections for IL-17A treatment studies

HEK293T cells were seeded 5 × 105cells per well (in 6-well plate) or 3 × 106 cells per 10-cm2 plates. The next day, 2 or 10 µg of plasmids was transfected using X-tremeGENE 9 DNA Transfection Reagent (Roche) in 6-well plate or 10-cm2 plates, respectively. For IL-17A (Millipore-Sigma, SRP0675) incubation in cells, 0.5 µg of IL-17A was treated either before or after transfection and incubated at 37°C. After 48 hours, cells were collected by trypsinization. For IL-17A incubation with cell lysates, transfected cell lysates were incubated in the presence of 0.5 or 5 µg/ml IL-17A at 4°C on a rotator overnight. Plasmids pLVX-EF1alpha–SARS-CoV-2–orf8-2xStrep-IRES-Puro (ORF8) and pLVX-EF1alpha-eGFP-2xStrep-IRES-Puro (EGFP-Strep) were a gift from N. J. Krogan. (Addgene plasmid #141390, 141395) (5). pLVX-EF1alpha- IRES-Puro (Vector) was obtained from Takara/Clontech.

SARS-CoV-2 ORF8 and IL17RA coimmunoprecipitation

Transfected and treated HEK293T cells were pelleted and washed in cold D-PBS and later resuspended in Flag-IP Buffer (50 mM Tris HCl, pH 7.4, with 150 mM NaCl, 1 mM EDTA, and 1% NP-40) with 1x HALT (ThermoFisher Scientific, 78429), incubated with buffer for 15 min on ice then centrifuged at 13,000 rpm for 5 min. The supernatant was collected and 1 mg of protein was used for immunoprecipitation (IP) with 100 µl of Streptactin Sepharose (IBA, 2-1201-010) on a rotor overnight at 4°C. Immunoprecipitates were washed five times with Flag-IP buffer and eluted with 1x Buffer E (100 mM Tris-Cl, 150 mM NaCl, 1 mM EDTA, 2.5 mM Desthiobiotin). Eluate was diluted with 1x-NuPAGE (ThermoFisher Scientific, #NP0008) LDS Sample Buffer with 2.5% β-

Mercaptoethanol and blotted for targeted antibodies. Antibodies used were Strep tag II (Qiagen, #34850), B-Actin (Sigma, #A5316), and IL17RA (Cell Signaling, #12661S).

Computational docking of PGES-2 and Nsp7

A model for human PGES-2 dimer was constructed by homology using MODELER (78) from the crystal structure of Macaca fascularis mPGES-2 [PDB ID: 1Z9H (79); 98% sequence identity] bound to indomethacin. Indomethacin was removed from the structure utilized for docking. The structure of SARS-CoV-2 Nsp7 was extracted from PDB ID 7BV2 (80). Docking models were produced using ClusPro (81), ZDock (82), HDock (83), Gramm-X (84), SwarmDock (85), and PatchDock (86) with SOAP-PP score (87). For each protocol, up to 100 top scoring models were extracted (fewer for those that do not report >100 models); for PatchDock, models with SOAP-PP z-scores >3.0 were used (fig. S23A). The 420 models were clustered at 4.0-Å RMSD, resulting in 127 clusters. The two largest clusters, composed of 192 models, are related by dimer symmetry. All other clusters contain <15 models.

Assessment of positive selection signatures in SIGMAR1

SIGMAR1 protein alignments were generated from whole-genome sequences of 359 mammals curated by the Zoonomia consortium. Protein alignments were generated with TOGA (https://github.com/hillerlab/TOGA), and missing sequence gaps were refined with CACTUS (88, 89). Branches undergoing positive selection were detected with the branch-site test aBSREL (90) implemented in the HyPhy package (90, 91). PhyloP was used to detect codons undergoing accelerated evolution along branches detected as undergoing positive selection by aBSREL relative to the neutral evolution rate in mammals, determined using phyloFit on third nucleotide positions of codons which are assumed to evolve neutrally. P values from phyloP were corrected for multiple tests using the Benjamini-Hochberg method (92). PhyloFit and phyloP are both part of the PHAST package version 1.4 (93, 94).

Comparative SARS-CoV-1 inhibition by amiodarone

SARS-CoV-1 (Urbani) drug screens were performed with Vero E6 cells (ATCC #1568, Manassas, VA) cultured in DMEM (Quality Biological), supplemented with 10% (v/v) heat-inactivated FBS (Sigma), 1% (v/v) penicillin-streptomycin (Gemini Bio-products), and 1% (v/v) L-glutamine (2 mM final concentration, Gibco). Cells were plated in opaque 96-well plates 1 day before infection. Drugs were diluted from stock to 50 μM and an 8-point 1:2 dilution series prepared in duplicate in Vero Media. Every compound dilution and control were normalized to contain the same

concentration of drug vehicle (e.g., DMSO). Cells were pretreated with drug for 2 hours at 37°C (5% CO2) before infection with SARS-CoV-1 at MOI 0.01. In addition to plates that were infected, parallel plates were left uninfected to monitor cytotoxicity of drug alone. All plates were incubated at 37°C (5% CO2) for 3 days before performing CellTiter-Glo (CTG) assays as per the manufacturer's instruction (Promega, Madison, WI). Luminescence was read on a BioTek Synergy HTX plate reader (BioTek Instruments Inc., Winooski, VT) using the Gen5 software (version 7.07, Biotek Instruments Inc., Winooski, VT).

Real-world data source and analysis

This study used deidentified patient-level records from HealthVerity's Marketplace dataset, a nationally representative dataset covering >300 million patients with medical and pharmacy records from >60 health care data sources in the United States. The current study used data from 738,933 patients with documented COVID-19 infection between 1 March 2020 and 17 August 2020, defined as a positive or presumptive positive viral laboratory test result or an International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) diagnosis code of U07.1 (COVID-19).

For this population, we analyzed medical claims, pharmacy claims, laboratory data, and hospital chargemaster data containing diagnoses, procedures, medications, and COVID-19 laboratory results from both inpatient and outpatient settings. Claims data included open (unadjudicated) claims sourced in near-real time from practice management and billing systems, claims clearinghouses, and laboratory chains, as well as closed (adjudicated) claims encompassing all major U.S. payer types (commercial, Medicare, and Medicaid). For inpatient treatment evaluations, we used linked hospital chargemaster data containing records of all billable procedures, medical services, and treatments administered in hospital settings. Linkage of patient-level records across these data types provides a longitudinal view of baseline health status, medication use, and COVID-19 progression for each patient under study. Data for this study covered the period of 1 December 2018 through 17 August 2020. All analyses were conducted with the Aetion Evidence Platform version r4.6.

This study was approved by the New England institutional review board (IRB) (no. 1-9757-1). Medical records constitute protected health information and can be made available to qualified individuals on reasonable request.

Observation of hospitalization outcomes in outpatient new users of indomethacin (treatment arm) versus celecoxib (active comparator) using real-world data

We used an incident (new) user, active comparator design (95, 96) to assess the risk of hospitalization among newly diagnosed COVID-19 patients who were subsequently treated with indomethacin or the comparator agent, celecoxib. Patients were required to have COVID-19 infection recorded in an outpatient setting during the study period of 1 March 2020 to 17 August 2020 and occurring in the 21 days before (and including) the date of indomethacin or celecoxib treatment initiation. Prevalent users of prescription-only NSAIDs (any prescription fill for indomethacin, celecoxib, ketoprofen, meloxicam, sulindac, or piroxicam 60 days prior) and patients hospitalized in the 21 days before and including the date of treatment initiation were excluded from this analysis.

Using RSS, patients treated with indomethacin were matched at a 1:1 ratio to controls randomly selected among patients treated with celecoxib, with direct matching on calendar date of treatment (±7 days), age (±5 years), sex, Charlson comorbidity index (exact) (97), time since confirmed COVID-19 (±5 days), and disease severity based on the highest-intensity COVID-19–related health service in the 7 days before and including the date of treatment initiation (laboratory service only versus outpatient medical visit versus emergency department visit) and symptom profile in the 21 days before and including the date of treatment initiation (recorded symptoms versus none). This risk-set–sampled population was further matched on a PS (33) estimated using logistic regression with 24 demographic and clinical risk factors, including covariates related to baseline medical history and COVID-19 severity in the 21 days before treatment (table S11). Balance between indomethacin and celecoxib treatment groups was evaluated by comparison of absolute standardized differences in covariates, with an absolute standardized difference of <0.2 indicating good balance between the treatment groups (98).

The primary analysis was an intention-to-treat design, with follow-up beginning 1 day after indomethacin or celecoxib initiation and ending on the earliest occurrence of 30 days of follow-up reached or end of patient data. ORs for the primary outcome of all-cause inpatient hospitalization were estimated for the RSS-plus-PS–matched population as well as for the RSS-matched population. Our primary outcome definition required a record of inpatient hospital admission with a resulting inpatient stay; as a sensitivity, a broader outcome definition captured any hospital visit (defined with revenue and place of service codes).

Observation of mechanical ventilation outcomes in inpatient new users of typical antipsychotics (treatment arm) versus atypical antipsychotics (active comparator) using real-world data

We used an incident user, active comparator design (95, 96) to assess the risk of mechanical ventilation among hospitalized COVID-19 patients treated with typical or atypical antipsychotics in an inpatient setting. See table S11 for a list of drugs included in each category. To permit assessment of day-level in-hospital confounders and outcomes, this analysis was restricted to hospitalized patients observable in hospital chargemaster data. Prevalent users of typical or atypical antipsychotics (any prescription fill or chargemaster-documented use in 60 days prior) and patients with evidence of mechanical ventilation in the 21 days before and including the date of treatment initiation were excluded from this analysis.

Using RSS, hospitalized patients treated with typical antipsychotics were matched at a 1:1 ratio to controls randomly selected among patients treated with atypical antipsychotics, with direct matching (1:1 fixed ratio) on calendar date of treatment (±7 days), age (±5 years), sex, Charlson comorbidity index (exact) (97), time since hospital admission, and disease severity as defined with a simplified version of the World Health Organization's ordinal scale for clinical improvement (99). This risk-set–sampled population was further matched on a PS estimated using logistic regression with 36 demographic and clinical risk factors, including covariates related to baseline medical history, admitting status, and disease severity at treatment (table S11). Balance between typical and atypical treatment groups was evaluated by comparison of absolute standardized differences in covariates, with an absolute standardized difference of <0.2 indicating good balance between the treatment groups (98).

The primary analysis was an intention-to-treat design, with follow up beginning 1 day after the date of typical or atypical antipsychotic treatment initiation and ending on the earliest occurrence of 30 days of follow-up reached, discharge from hospital, or end of patient data. ORs for the primary outcome of inpatient mechanical ventilation were estimated for the RSS-plus-PS–matched population as well as for the RSS-matched population.

Fig. 1 Coronavirus genome annotations and integrative analysis overview.

(**A**) Genome annotation of SARS-CoV-2, SARS-CoV-1, and MERS-CoV with putative protein coding genes highlighted. Intensity of filled color indicates the lowest sequence identity between SARS-CoV-2 and SARS-CoV-1 or between SARS-CoV-2 and MERS. (**B** to **D**) Genome annotation of structural protein genes for SARS-CoV-2 (B), SARS-CoV-1 (C), and MERS-CoV (**D**). Color intensity indicates sequence identity to specified virus. (**E**) Overview of comparative coronavirus analysis. Proteins from SARS-CoV-2, SARS-CoV-1, and MERS-CoV were analyzed for their protein interactions and subcellular localization, and these data were integrated for comparative host interaction network analysis, followed by functional, structural, and clinical data analyses for exemplary virus-specific and pan-viral interactions. The asterisk indicates that the SARS-CoV-2 interactome was previously published in a separate study (5). SARS, both SARS-CoV-1 and SARS-CoV-2; MERS, MERS-CoV; Nsp, nonstructural protein; ORF, open reading frame.

Fig. 2 Coronavirus protein localization analysis.

(**A**) Overview of experimental design to determine localization of Strep-tagged SARS-CoV-2, SARS-CoV-1, and MERS-CoV proteins in HeLaM cells (left) or of viral proteins upon SARS-CoV-2 infection in Caco-2 cells (right). (**B**) Relative localization for all coronavirus proteins across viruses expressed individually (blue color bar) or in SARS-CoV-2–infected cells (colored box outlines). (**C** and **D**) Localization of Nsp1 and ORF3a expressed individually (C) or during infection (**D**); for representative images of all tagged constructs and viral proteins imaged during infection, see figs. S8 to S14 and fig. S15, respectively. Scale bars, 10 μm. (**E**) Prey overlap per bait measured as Jaccard index comparing SARS-CoV-2 versus SARS-CoV-1 (red dots) and SARS-CoV-2 versus MERS-CoV (blue dots) for all viral baits (all), viral baits found in the same cellular compartment (yes), and viral baits found in different compartments (no).

Fig. 3 Comparative analysis of coronavirus-host interactomes.

(**A**) Clustering analysis (K-means) of interactors from SARS-CoV-2, SARS-CoV-1, and MERS-CoV, weighted according to the average between their MiST and SAINT scores (interaction score K). Included are only viral protein baits represented amongst all three viruses and interactions that pass the high-confidence scoring threshold for at least one virus. Seven clusters highlight all possible scenarios of shared versus individual interactions, and percentages of total interactions are noted. (**B**) GO enrichment analysis of each cluster from (A), with the top six most-significant terms per cluster. Color indicates −log10(q), and the number of genes with significant ($q < 0.05$;

white) or nonsignificant enrichment ($q > 0.05$; gray) is shown. (**C**) Percentage of interactions for each viral protein belonging to each cluster identified in (A). (**D**) Correlation between protein sequence identity and PPI overlap (Jaccard index) comparing SARS-CoV-2 and SARS-CoV-1 (blue) or MERS-CoV (red). Interactions for PPI overlap are derived from the final thresholded list of interactions per virus. (**E**) GO biological process terms significantly enriched ($q < 0.05$) for all three virus PPIs with Jaccard index indicating overlap of genes from each term for pairwise comparisons between SARS-CoV-1 and SARS-CoV-2 (purple), SARS-CoV-1 and MERS-CoV (green), and SARS-CoV-2 and MERS-CoV (orange). (**F**) Fraction of shared preys between orthologous (blue) and nonorthologous (red) viral protein baits. (**G**) Heatmap depicting overlap in PPIs (Jaccard index) between each bait from SARS-CoV-2 and MERS-CoV. Baits in gray were not assessed, do not exist, or do not have high-confidence interactors in the compared virus. Nonorthologous bait interactions are highlighted with a red square. GO, gene ontology; PPI, protein-protein interaction; SARS2, SARS-CoV-2; SARS1, SARS-CoV-1; MERS, MERS-CoV.

Fig. 4 Comparative differential interaction analysis reveals shared virus-host interactions.

(**A**) Flowchart depicting calculation of DIS values using the average between the SAINT and MiST scores between every bait (i) and prey (j) to derive interaction score (K). The DIS is the difference between the interaction scores from each virus. The modified DIS (SARS-MERS) compares the average K from SARS-CoV-1 and SARS-CoV-2 to that of MERS-CoV (see Materials and methods). Only viral bait proteins shared between all three viruses are included. (**B**) Density histogram of the DISs for all comparisons. (**C**) Dot plot depicting the DISs of interactions from viral bait proteins shared between all three viruses, ordered left to right by the mean DIS per viral bait. (**D**) Virus-human PPI map depicting the SARS-MERS comparison [purple in (B) and (C)]. The network depicts interactions derived from cluster 2 (all three viruses), cluster 4 (SARS-CoV-1 and SARS-CoV-2), and cluster 5 (MERS-CoV only). Edge color denotes DIS: red indicates interactions specific to SARS-CoV-1 and SARS-CoV-2 but absent in MERS-CoV; blue indicates interactions specific to MERS-CoV but absent from both SARS-CoV-1 and SARS-CoV-2; and black indicates interactions shared between all three viruses. Human-human interactions (thin dark gray line) and proteins sharing the same protein complexes or biological processes (light yellow or light blue highlighting, respectively) are shown. Host-host physical interactions, protein complex definitions, and biological process groupings are derived from CORUM (46), GO (biological process), and manually curated from literature sources. Thin dashed gray lines are used to indicate the placement of node labels when adjacent node labels would have otherwise

been obscured. DIS, differential interaction score; SARS2, SARS-CoV-2; SARS1, SARS-CoV-1; MERS, MERS-CoV; SARS, both SARS-CoV-1 and SARS-CoV-2.

Fig. 5 Functional interrogation of SARS-CoV-2 interactors using genetic perturbations.

(**A**) A549-ACE2 cells were transfected with siRNA pools targeting each of the human genes from the SARS-CoV-2 interactome, followed by infection with SARS-CoV-2 and virus quantification using RT-qPCR. Cell viability and knockdown efficiency in uninfected cells was determined in parallel. (**B**) Caco-2 cells with CRISPR knockouts (KO) of each human gene from the SARS-CoV-2 interactome were infected with SARS-CoV-2, and supernatants were serially diluted and plated onto Vero E6 cells for quantification. Viabilities of the uninfected CRISPR knockout cells after infection were determined in parallel by DAPI staining. (**C** and **D**) Plot of results from the infectivity screens in A549-ACE2 knockdown cells (C) and Caco-2 knockout cells (D) sorted by z-score ($z < 0$, decreased infectivity; $z > 0$ increased infectivity). Negative controls (nontargeting control for siRNA, nontargeted cells for CRISPR) and positive controls (ACE2 knockdown or knockout) are highlighted. (**E**) Results from both assays with potential hits ($|z|>2$) highlighted in red (A549-ACE2), yellow (Caco-2), and orange (both). (**F**) Pan-coronavirus interactome reduced to human preys with significant increase (red nodes) or decrease (blue nodes) in SARS-CoV2 replication upon knockdown or knockout. Viral proteins baits from SARS-CoV-2 (red), SARS-CoV-1 (orange), and MERS-CoV (yellow) are represented as diamonds. The thickness of the edge indicates the strength of the PPI in spectral counts. KD, knockdown; KO, knockout; PPI, protein-protein interaction.

Fig. 6 Interaction between ORF9b and human Tom70.

(**A**) ORF9b-Tom70 interaction is conserved between SARS-CoV-1 and SARS-CoV-2. (**B**) Viral titers in Caco-2 cells after CRISPR knockout of TOMM70 or controls. (**C**) Coimmunoprecipitation of endogenous Tom70 with Strep-tagged ORF9b from SARS-CoV-1 and SARS-CoV-2; Nsp2 from SARS-CoV-1, SARS-CoV-2, and MERS-CoV; or vector control in HEK293T cells. Representative blots of whole-cell lysates and eluates after IP are shown. (**D**) Size exclusion chromatography traces (10/300 S200 increase) of ORF9b alone, Tom70 alone, and coexpressed ORF9b-Tom70 complex purified from recombinant expression in E. coli. Insert shows SDS-PAGE of the complex peak indicating presence of both proteins. (**E**) Immunostainings for Tom70 in HeLaM cells transfected with GFP-Strep and ORF9b from SARS-CoV-1 and SARS-CoV-2 (left) and mean fluorescence intensity ± SD values of Tom70 in GFP-Strep and ORF9b expressing cells (normalized to nontransfected cells) (right). Scale bar, 10 μm. (**F**) Flag-Tom70 expression

levels in total cell lysates of HEK293T cells upon titration of cotransfected Strep-ORF9b from SARS-CoV-1 and SARS-CoV-2. (**G**) Immunostaining for ORF9b and Tom70 in Caco-2 cells infected with SARS-CoV-2 (left) and mean fluorescence intensity ± SD values of Tom70 in uninfected and SARS-CoV-2–infected cells (right). SARS2, SARS-CoV-2; SARS1, SARS-CoV-1; MERS, MERS-CoV; IP, immunoprecipitation. **P < 0.05, Student's t test.

Fig. 7 Cryo-EM structure of ORF9b-Tom70 complex reveals ORF9b adopting a helical fold and binding at the substrate recognition site of Tom70.

(**A**) Surface representation of the ORF9b-Tom70 structure. Tom70 is depicted as molecular surface in green, ORF9b is depicted as ribbon in orange. Region in charcoal indicates Hsp70 or Hsp90 binding site on Tom70. (**B**) Magnified view of ORF9b-Tom70 interactions with interacting hydrophobic residues on Tom70 indicated and shown in spheres. The two phosphorylation sites on ORF9b, S50 and S53, are shown in yellow. (**C**) Ionic interactions between Tom70 and ORF9b are depicted as sticks. Highly conserved residues on Tom70 making hydrophobic interactions with ORF9b are depicted as spheres. (**D**) Diagram depicting secondary structure comparison of ORF9b as predicted by JPred server—as visualized in our structure—or as visualized in the previously crystallized dimer structure (PDB ID: 6Z4U) (16). Pink tubes indicate helices, charcoal arrows indicate beta strands, and the amino acid sequence for the region visualized in the cryo-EM structure is shown on top. (**E**) Predicted probability of having an internal MTS as output by TargetP server by serially running N-terminally truncated regions of SARS-CoV-2 ORF9b. Region visualized in the cryo-EM structure (amino acids 39 to 76) overlaps with the highest internal MTS probability region (amino acids 40 to 50). MTS, mitochondrial targeting signal. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

Fig. 8 SARS-CoV-2 ORF8 and functional interactor IL17RA are linked to viral outcomes.

(**A**) IL17RA and ADAM9 are functional interactors of SARS-CoV-2 ORF8. Only interactors identified in the genetic screening are shown. (**B**) Coimmunoprecipitation of endogenous IL17RA with Strep-tagged ORF8 or EGFP with or without IL-17A treatment at different times. Overexpression was done in HEK293T cells. (**C**) Viral titer after IL17RA or control knockdown in A549-ACE2 cells. (**D**) OR of membership in indicated cohorts by genetically predicted sIL17RA levels. SARS2, SARS-CoV-2; IP, immunoprecipitation; SD, standard deviation; OR, odds ratio;

CI, confidence interval; sIL17RA, soluble IL17RA. *P < 0.05, unpaired t test. Error bars in (C) indicate SDs; in (D), they indicate 95% CIs.

Fig. 9 Real-world data analysis of drugs identified through molecular investigation support their antiviral activity.

(**A**) Schematic of retrospective real-world clinical data analysis of indomethacin use for outpatients with SARS-CoV-2. Plots show distribution of propensity scores (PSs) for all included patients (red, indomethacin users; blue, celecoxib users). For a full list of inclusion, exclusion, and matching criteria, see Materials and methods and table S11. (**B**) Effectiveness of indomethacin versus celecoxib in patients with confirmed SARS-CoV-2 infection treated in an outpatient setting. Average standardized absolute mean difference (ASAMD) is a measure of balance between indomethacin and celecoxib groups calculated as the mean of the absolute standardized difference for each PS factor (table S11); P value and ORs with 95% CIs are estimated using the Aetion Evidence Platform r4.6. No ASAMD was >0.1. (**C**) Schematic of retrospective real-world clinical data analysis of typical antipsychotic use for inpatients with SARS-CoV-2. Plots show distribution of PSs for all included patients (red, typical users; blue, atypical users). For a full list of inclusion, exclusion, and matching criteria see Materials and methods and table S11. (**D**) Effectiveness of typical versus atypical antipsychotics among hospitalized patients with confirmed SARS-CoV-2 infection treated in hospital. ASAMD is a measure of balance between typical and atypical groups calculated as the mean of the absolute standardized difference for each PS factor (table S11); P value and ORs with 95% CIs are estimated using the Aetion Evidence Platform r4.6. No ASAMD was >0.1.

Figure 1

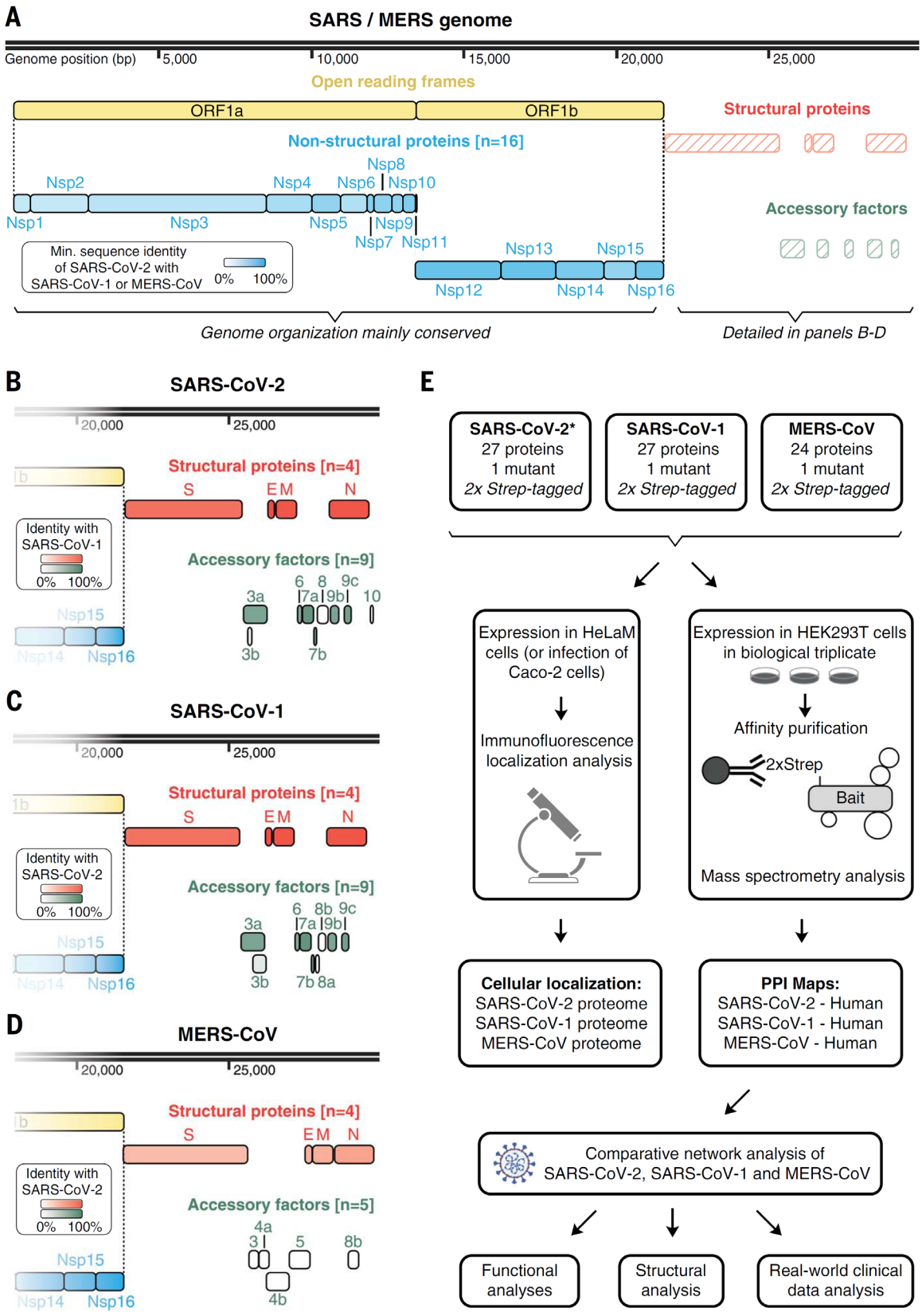

A
SARS / MERS genome
Genome position (bp)
5,000
10,000
15,000
20,000
25,000
Open reading frames
ORF1a
ORF1b
Structural proteins
Non-structural proteins [n=16]
Nsp1
Nsp2
Nsp3
Nsp4
Nsp5
Nsp6
Nsp7
Nsp8
Nsp9
Nsp10
Nsp11
Nsp12
Nsp13
Nsp14
Nsp15
Nsp16
Accessory factors
Min. sequence identity of SARS-CoV-2 with SARS-CoV-1 or MERS-CoV
0%
100%
Genome organization mainly conserved
Detailed in panels B-D
B
SARS-CoV-2
Structural proteins [n=4]
S
E
M
N
Identity with SARS-CoV-1
Accessory factors [n=9]
3a
3b
6
7a
7b
8
9b
9c
10
C
SARS-CoV-1
Identity with SARS-CoV-2
Accessory factors [n=9]
8a
D
MERS-CoV
Accessory factors [n=5]
3
4a
4b
5
8b
E
SARS-CoV-2*
27 proteins
1 mutant
2x Strep-tagged
SARS-CoV-1
MERS-CoV
24 proteins
Expression in HeLaM cells (or infection of Caco-2 cells)
Immunofluorescence localization analysis
Expression in HEK293T cells in biological triplicate
Affinity purification
2xStrep
Bait
Mass spectrometry analysis
Cellular localization:
SARS-CoV-2 proteome
SARS-CoV-1 proteome
MERS-CoV proteome
PPI Maps:
SARS-CoV-2 - Human
SARS-CoV-1 - Human
MERS-CoV - Human
Comparative network analysis of SARS-CoV-2, SARS-CoV-1 and MERS-CoV
Functional analyses
Structural analysis
Real-world clinical data analysis

Figure 2

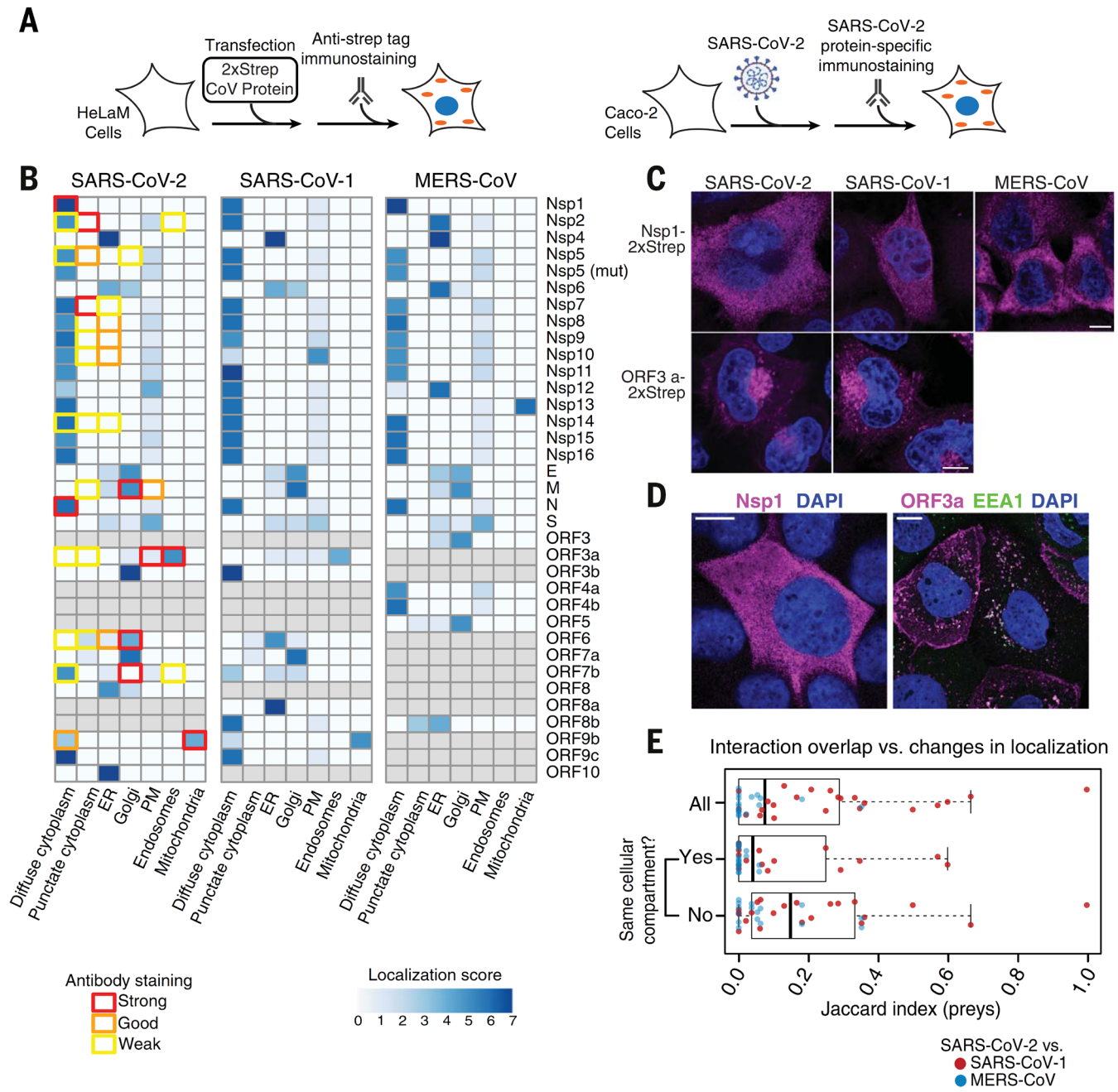

A
Transfection
2xStrep
CoV Protein
Anti-strep tag
immunostaining
HeLaM
Cells
SARS-CoV-2
SARS-CoV-2
protein-specific
immunostaining
Caco-2
Cells
B
SARS-CoV-2
SARS-CoV-1
MERS-CoV
Nsp1
Nsp2
Nsp4
Nsp5
Nsp5 (mut)
Nsp6
Nsp7
Nsp8
Nsp9
Nsp10
Nsp11
Nsp12
Nsp13
Nsp14
Nsp15
Nsp16
E
M
N
S
ORF3
ORF3a
ORF3b
ORF4a
ORF4b
ORF5
ORF6
ORF7a
ORF7b
ORF8
ORF8a
ORF8b
ORF9b
ORF9c
ORF10
Diffuse cytoplasm
Punctate cytoplasm
ER
Golgi
PM
Endosomes
Mitochondria
Antibody staining
Strong
Good
Weak
Localization score
0 1 2 3 4 5 6 7
C
SARS-CoV-2
SARS-CoV-1
MERS-CoV
Nsp1-
2xStrep
ORF3 a-
2xStrep
D
Nsp1 DAPI
ORF3a EEA1 DAPI
E
Interaction overlap vs. changes in localization
All
Yes
No
Same cellular
compartment?
0.0
0.2
0.4
0.6
0.8
1.0
Jaccard index (preys)
SARS-CoV-2 vs.
SARS-CoV-1
MERS-CoV

Figure 3

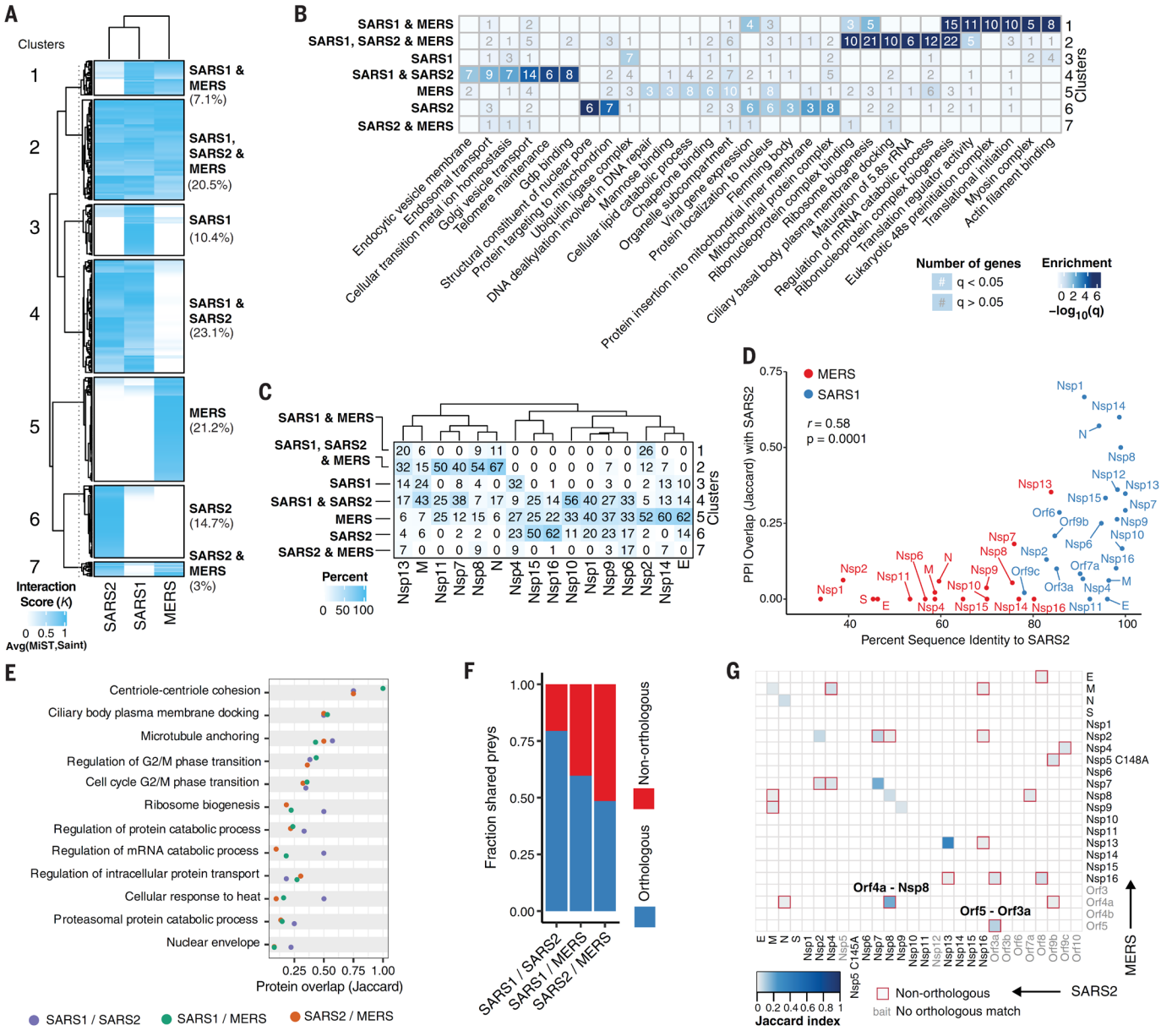

A
Clusters
SARS1 & MERS (7.1%)
SARS1, SARS2 & MERS (20.5%)
SARS1 (10.4%)
SARS1 & SARS2 (23.1%)
MERS (21.2%)
SARS2 (14.7%)
SARS2 & MERS (3%)
Interaction Score (K)
Avg(MiST,Saint)
SARS2
SARS1
MERS
B
SARS1 & MERS
SARS1, SARS2 & MERS
SARS1
SARS1 & SARS2
MERS
SARS2
SARS2 & MERS
Clusters
Endocytic vesicle membrane
Endosomal transport
Cellular transition metal ion homeostasis
Golgi vesicle transport
Telomere maintenance
Gdp binding
Structural constituent of nuclear pore
Protein targeting to mitochondrion
DNA dealkylation involved in DNA repair
Ubiquitin ligase complex
Cellular lipid catabolic process
Mannose binding
Chaperone binding
Organelle subcompartment
Viral gene expression
Protein localization to nucleus
Flemming body
Protein insertion into mitochondrial inner membrane
Mitochondrial protein complex
Ribonucleoprotein complex binding
Ciliary basal body plasma membrane docking
Ribosome biogenesis
Maturation of 5.8s rRNA
Regulation of mRNA catabolic process
Ribonucleoprotein complex biogenesis
Eukaryotic 48s preinitiation complex
Translation regulator activity
Translational initiation
Myosin complex
Actin filament binding
Number of genes
q < 0.05
q > 0.05
Enrichment
−log10(q)
C
SARS1 & MERS
SARS1, SARS2 & MERS
SARS1
SARS1 & SARS2
MERS
SARS2
SARS2 & MERS
Clusters
Nsp13 M Nsp11 Nsp7 Nsp8 N Nsp4 Nsp15 Nsp16 Nsp10 Nsp1 Nsp9 Nsp6 Nsp2 Nsp14 E
Percent
D
MERS
SARS1
r = 0.58
p = 0.0001
PPI Overlap (Jaccard) with SARS2
Percent Sequence Identity to SARS2
E
Centriole-centriole cohesion
Ciliary body plasma membrane docking
Microtubule anchoring
Regulation of G2/M phase transition
Cell cycle G2/M phase transition
Ribosome biogenesis
Regulation of protein catabolic process
Regulation of mRNA catabolic process
Regulation of intracellular protein transport
Cellular response to heat
Proteasomal protein catabolic process
Nuclear envelope
Protein overlap (Jaccard)
SARS1 / SARS2
SARS1 / MERS
SARS2 / MERS
F
Fraction shared preys
Non-orthologous
Orthologous
SARS1 / SARS2
SARS1 / MERS
SARS2 / MERS
G
Orf4a - Nsp8
Orf5 - Orf3a
MERS
SARS2
Jaccard index
Non-orthologous
bait No orthologous match

Figure 4

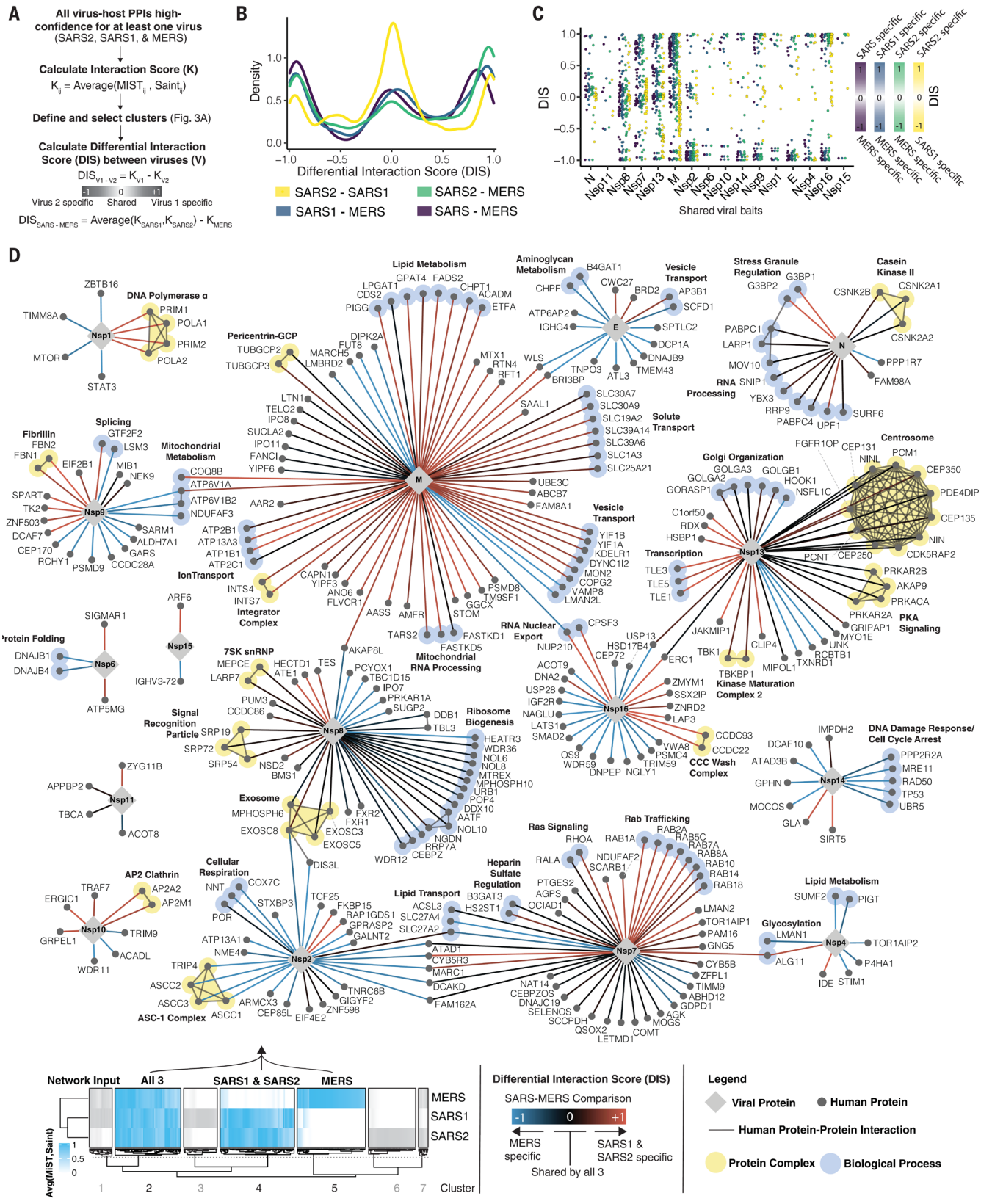

A
All virus-host PPIs high-confidence for at least one virus (SARS2, SARS1, & MERS)
Calculate Interaction Score (K)
K_ij = Average(MIST_ij , Saint_ij)
Define and select clusters (Fig. 3A)
Calculate Differential Interaction Score (DIS) between viruses (V)
DIS_V1 - V2 = K_V1 - K_V2
-1 0 +1
Virus 2 specific Shared Virus 1 specific
DIS_SARS - MERS = Average(K_SARS1,K_SARS2) - K_MERS
B
Density
Differential Interaction Score (DIS)
SARS2 - SARS1
SARS2 - MERS
SARS1 - MERS
SARS - MERS
C
DIS
SARS specific
SARS1 specific
SARS2 specific
SARS2 specific
MERS specific
MERS specific
MERS specific
SARS1 specific
N Nsp11 Nsp8 Nsp7 Nsp13 M Nsp2 Nsp6 Nsp10 Nsp14 Nsp9 Nsp1 E Nsp4 Nsp16 Nsp15
Shared viral baits
D
Nsp1
DNA Polymerase α
Nsp9
Fibrillin
Splicing
Mitochondrial Metabolism
IonTransport
Integrator Complex
M
Lipid Metabolism
Pericentrin-GCP
Vesicle Transport
Solute Transport
Nsp15
Nsp6
Protein Folding
Nsp8
7SK snRNP
Signal Recognition Particle
Mitochondrial RNA Processing
Ribosome Biogenesis
Exosome
Nsp11
Nsp2
Cellular Respiration
AP2 Clathrin
ASC-1 Complex
Lipid Transport
Nsp10
Nsp16
RNA Nuclear Export
CCC Wash Complex
Nsp7
Ras Signaling
Rab Trafficking
Heparin Sulfate Regulation
Aminoglycan Metabolism
Nsp13
Transcription
Kinase Maturation Complex 2
Golgi Organization
Centrosome
PKA Signaling
N
Stress Granule Regulation
Casein Kinase II
RNA Processing
Nsp14
DNA Damage Response/ Cell Cycle Arrest
Nsp4
Glycosylation
Lipid Metabolism
Network Input
All 3
SARS1 & SARS2
MERS
MERS
SARS1
SARS2
Avg(MIST,Saint)
Cluster
Differential Interaction Score (DIS)
SARS-MERS Comparison
-1 0 +1
MERS specific
SARS1 & SARS2 specific
Shared by all 3
Legend
Viral Protein
Human Protein
Human Protein-Protein Interaction
Protein Complex
Biological Process

Figure 5

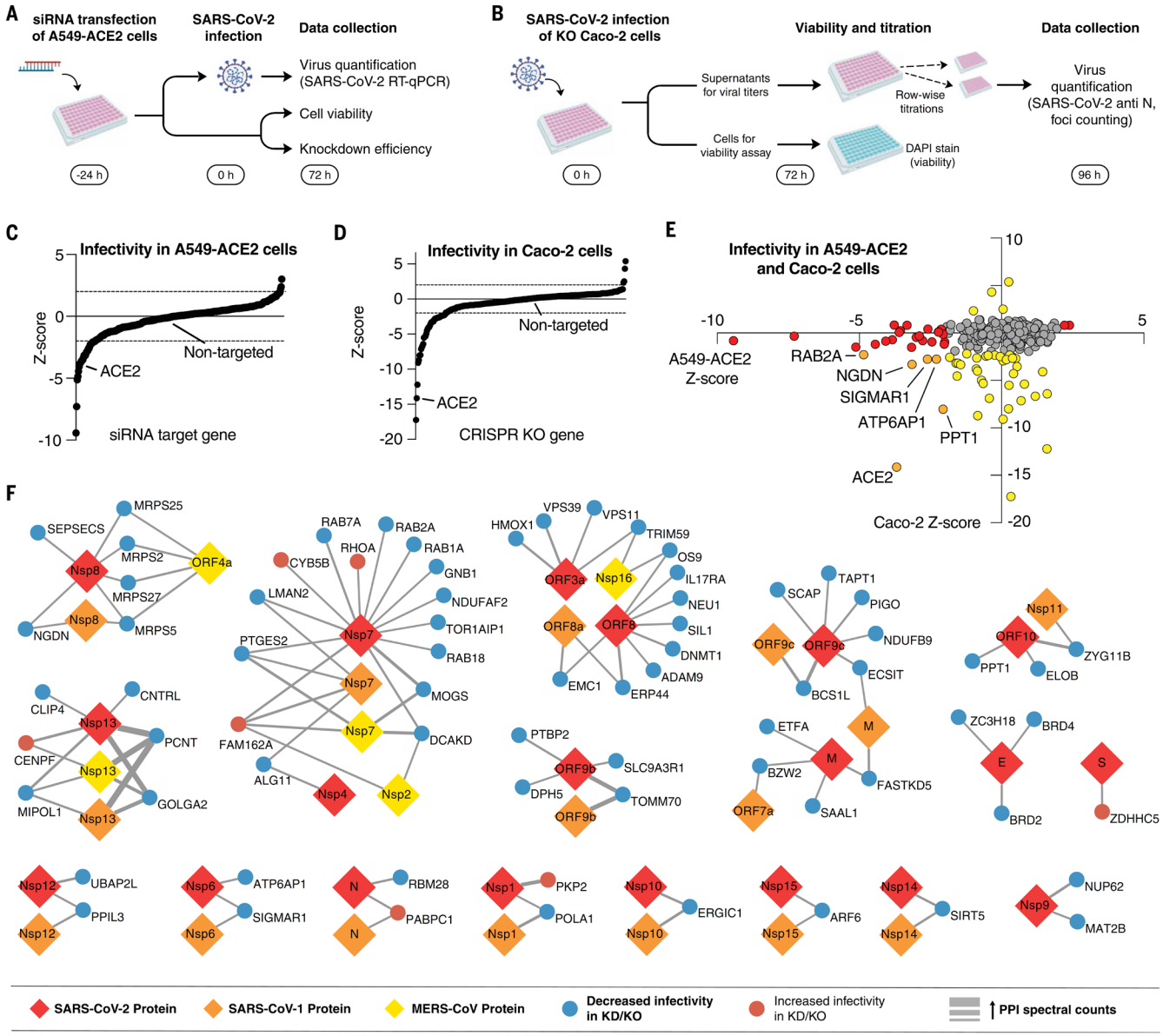

A
siRNA transfection of A549-ACE2 cells
SARS-CoV-2 infection
Data collection
Virus quantification (SARS-CoV-2 RT-qPCR)
Cell viability
Knockdown efficiency
-24 h
0 h
72 h
B
SARS-CoV-2 infection of KO Caco-2 cells
Viability and titration
Data collection
Supernatants for viral titers
Cells for viability assay
Row-wise titrations
DAPI stain (viability)
Virus quantification (SARS-CoV-2 anti N, foci counting)
0 h
72 h
96 h
C
Infectivity in A549-ACE2 cells
Z-score
Non-targeted
ACE2
siRNA target gene
D
Infectivity in Caco-2 cells
Z-score
Non-targeted
ACE2
CRISPR KO gene
E
Infectivity in A549-ACE2 and Caco-2 cells
A549-ACE2 Z-score
Caco-2 Z-score
RAB2A
NGDN
SIGMAR1
ATP6AP1
PPT1
ACE2
F
MRPS25
SEPSECS
MRPS2
ORF4a
Nsp8
MRPS27
NGDN
MRPS5
CLIP4
CNTRL
Nsp13
PCNT
CENPF
MIPOL1
GOLGA2
RAB7A
RAB2A
RHOA
RAB1A
CYB5B
GNB1
LMAN2
NDUFAF2
PTGES2
TOR1AIP1
Nsp7
RAB18
MOGS
FAM162A
DCAKD
ALG11
Nsp4
Nsp2
HMOX1
VPS39
VPS11
TRIM59
OS9
ORF3a
Nsp16
IL17RA
NEU1
ORF8a
ORF8
SIL1
DNMT1
EMC1
ADAM9
ERP44
PTBP2
ORF9b
SLC9A3R1
DPH5
TOMM70
TAPT1
SCAP
PIGO
ORF9c
NDUFB9
ECSIT
BCS1L
ETFA
M
BZW2
FASTKD5
ORF7a
SAAL1
Nsp11
ORF10
ZYG11B
PPT1
ELOB
ZC3H18
BRD4
E
S
BRD2
ZDHHC5
Nsp12
UBAP2L
PPIL3
Nsp6
ATP6AP1
SIGMAR1
N
RBM28
PABPC1
Nsp1
PKP2
POLA1
Nsp10
ERGIC1
Nsp15
ARF6
Nsp14
SIRT5
Nsp9
NUP62
MAT2B
SARS-CoV-2 Protein
SARS-CoV-1 Protein
MERS-CoV Protein
Decreased infectivity in KD/KO
Increased infectivity in KD/KO
PPI spectral counts

Figure 6

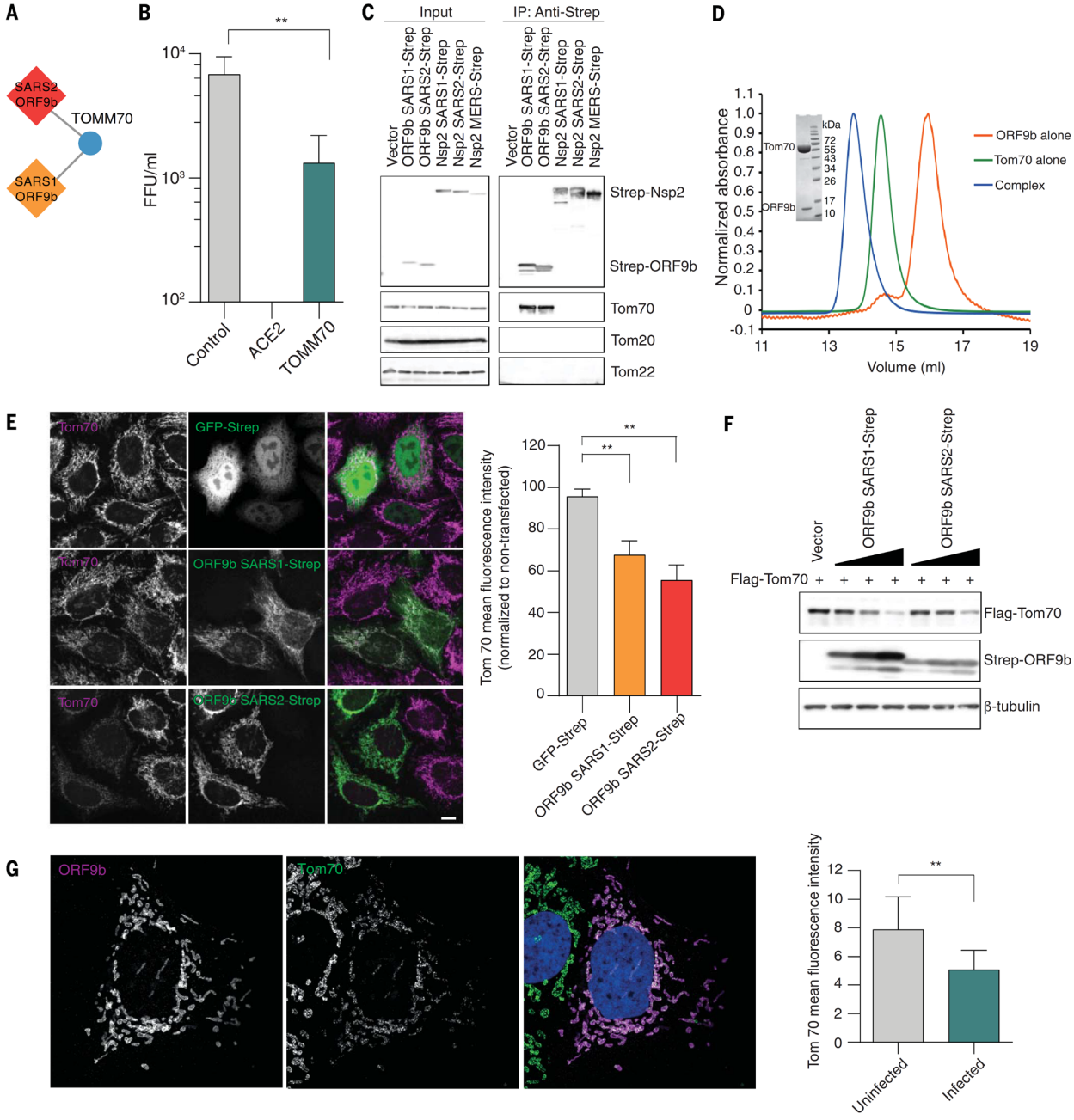

A
SARS2 ORF9b
TOMM70
SARS1 ORF9b
B
**
FFU/ml
Control
ACE2
TOMM70
C
Input
IP: Anti-Strep
Vector
ORF9b SARS1-Strep
ORF9b SARS2-Strep
Nsp2 SARS1-Strep
Nsp2 SARS2-Strep
Nsp2 MERS-Strep
Strep-Nsp2
Strep-ORF9b
Tom70
Tom20
Tom22
D
Normalized absorbance
Volume (ml)
kDa
ORF9b alone
Tom70 alone
Complex
E
Tom70
GFP-Strep
ORF9b SARS1-Strep
ORF9b SARS2-Strep
Tom 70 mean fluorescence intensity (normalized to non-transfected)
F
Vector
ORF9b SARS1-Strep
ORF9b SARS2-Strep
Flag-Tom70
Strep-ORF9b
β-tubulin
G
ORF9b
Tom70
Tom 70 mean fluorescence intensity
Uninfected
Infected

Figure 7

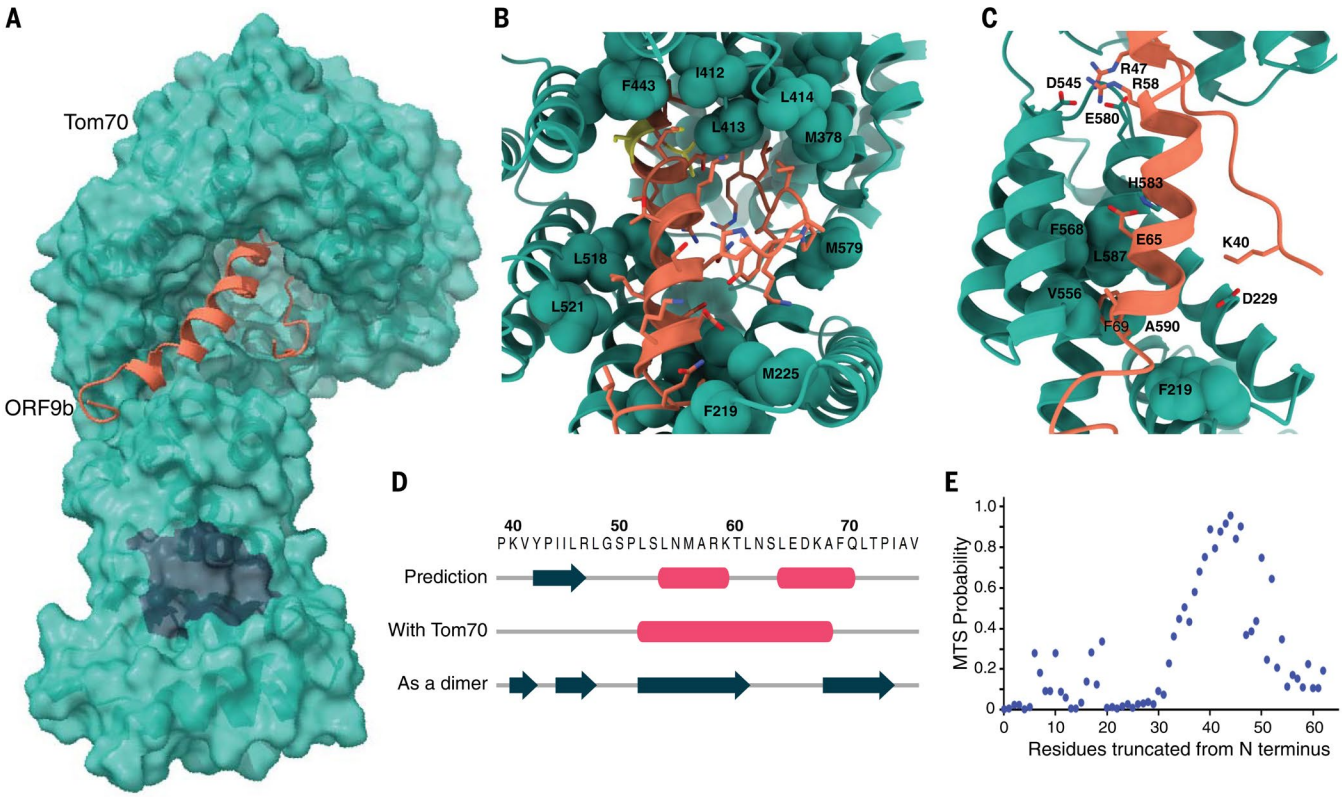

A
Tom70
ORF9b
B
F443
I412
L414
L413
M378
L518
M579
L521
M225
F219
C
R47
D545
R58
E580
H583
F568
E65
K40
L587
V556
D229
F69
A590
F219
D
40
50
60
70
PKVYPIILRLGSPLSLNMARKTLNSLEDKAFQLTPIAV
Prediction
With Tom70
As a dimer
E
MTS Probability
1.0
0.8
0.6
0.4
0.2
0
0
10
20
30
40
50
60
Residues truncated from N terminus

Figure 8

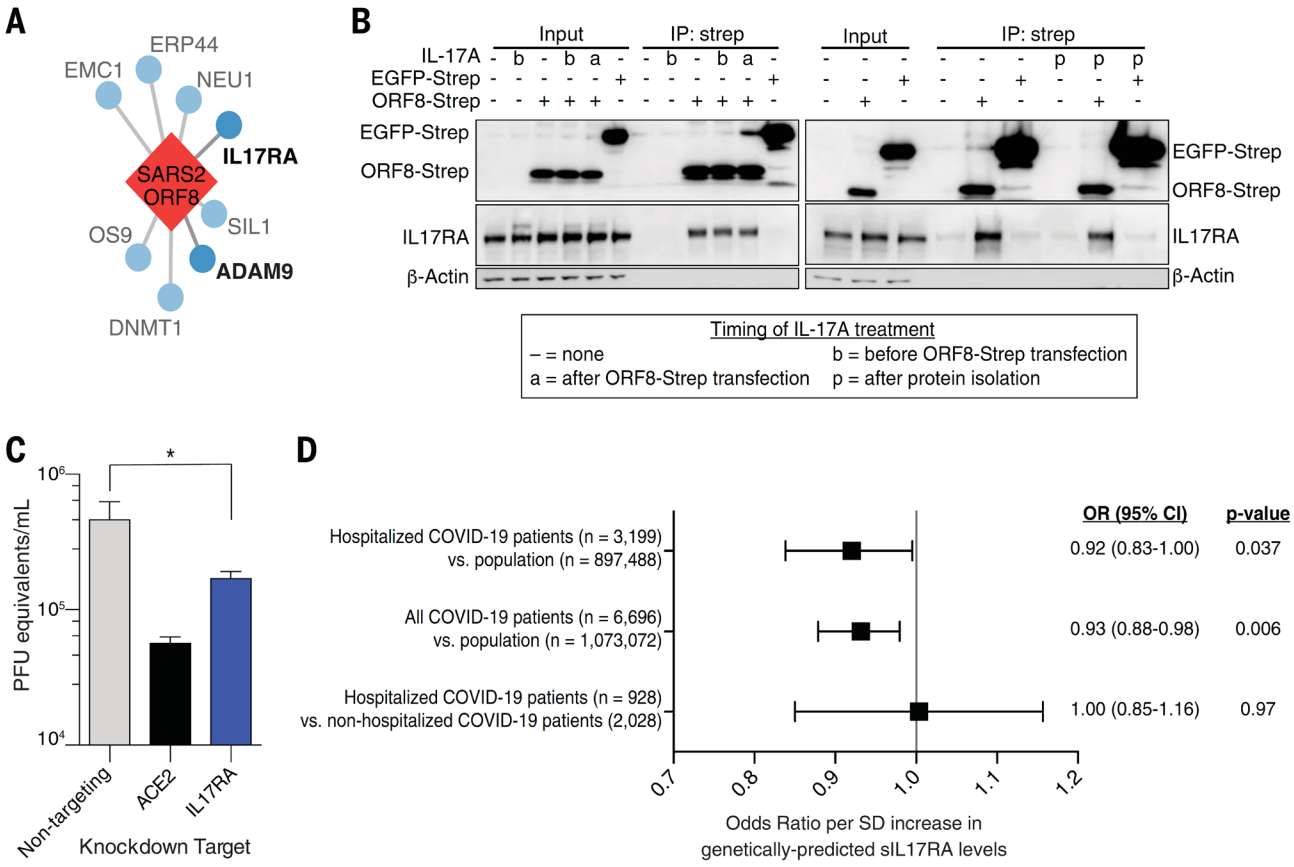
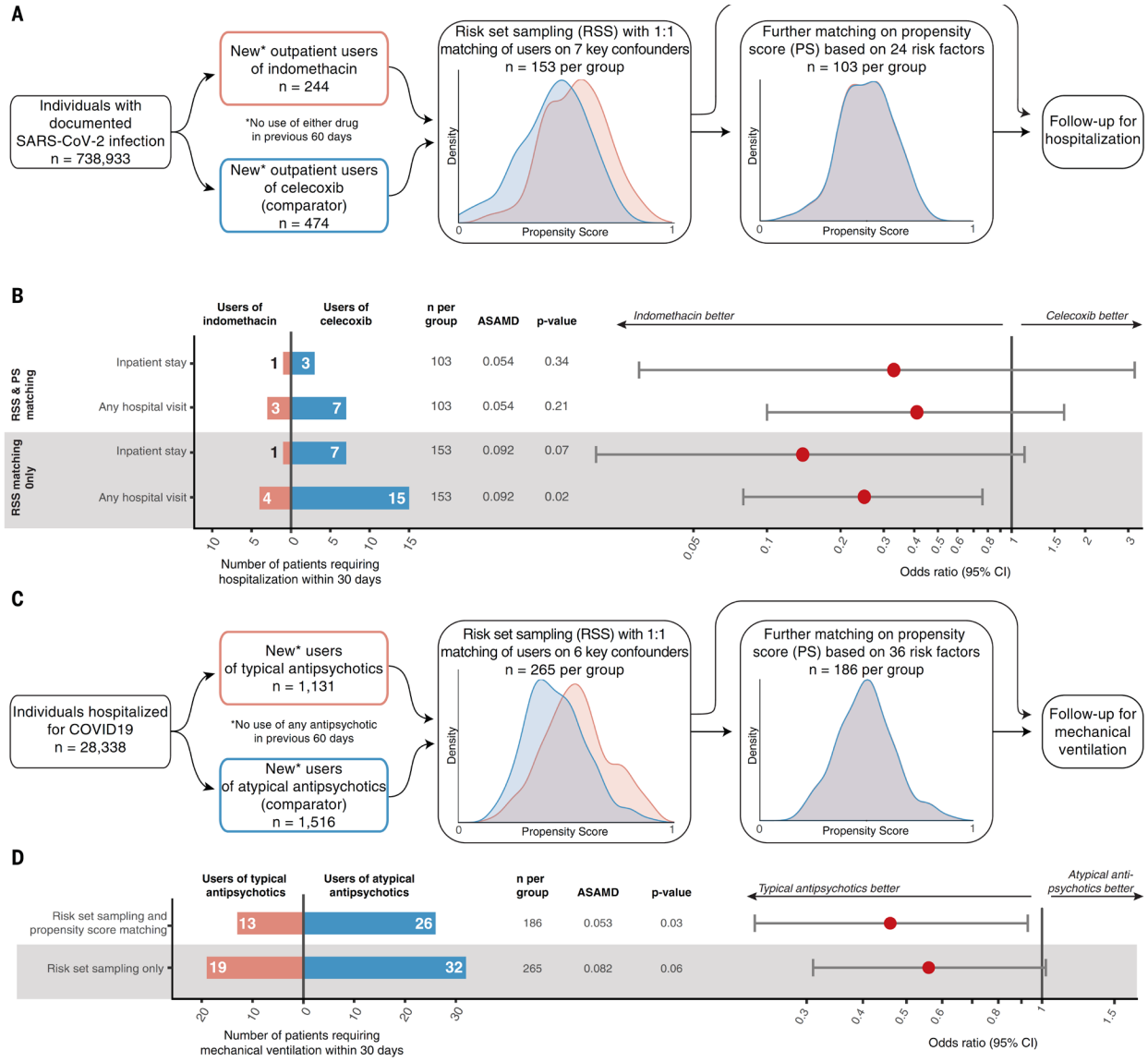

A
ERP44
EMC1
NEU1
IL17RA
SARS2 ORF8
SIL1
OS9
ADAM9
DNMT1
B
Input
IP: strep
Input
IP: strep
IL-17A
EGFP-Strep
ORF8-Strep
EGFP-Strep
ORF8-Strep
IL17RA
β-Actin
Timing of IL-17A treatment
– = none
b = before ORF8-Strep transfection
a = after ORF8-Strep transfection
p = after protein isolation
C
PFU equivalents/mL
Non-targeting
ACE2
IL17RA
Knockdown Target
D
OR (95% CI)
p-value
Hospitalized COVID-19 patients (n = 3,199) vs. population (n = 897,488)
0.92 (0.83-1.00)
0.037
All COVID-19 patients (n = 6,696) vs. population (n = 1,073,072)
0.93 (0.88-0.98)
0.006
Hospitalized COVID-19 patients (n = 928) vs. non-hospitalized COVID-19 patients (2,028)
1.00 (0.85-1.16)
0.97
Odds Ratio per SD increase in genetically-predicted sIL17RA levels

Figure 9


A
Individuals with documented SARS-CoV-2 infection n = 738,933
New* outpatient users of indomethacin n = 244
*No use of either drug in previous 60 days
New* outpatient users of celecoxib (comparator) n = 474
Risk set sampling (RSS) with 1:1 matching of users on 7 key confounders n = 153 per group
Further matching on propensity score (PS) based on 24 risk factors n = 103 per group
Follow-up for hospitalization
Density
Propensity Score
B
Users of indomethacin
Users of celecoxib
n per group
ASAMD
p-value
Indomethacin better
Celecoxib better
RSS & PS matching
Inpatient stay 1 3 103 0.054 0.34
Any hospital visit 3 7 103 0.054 0.21
RSS matching Only
Inpatient stay 1 7 153 0.092 0.07
Any hospital visit 4 15 153 0.092 0.02
Number of patients requiring hospitalization within 30 days
Odds ratio (95% CI)
C
Individuals hospitalized for COVID19 n = 28,338
New* users of typical antipsychotics n = 1,131
*No use of any antipsychotic in previous 60 days
New* users of atypical antipsychotics (comparator) n = 1,516
Risk set sampling (RSS) with 1:1 matching of users on 6 key confounders n = 265 per group
Further matching on propensity score (PS) based on 36 risk factors n = 186 per group
Follow-up for mechanical ventilation
Density
Propensity Score
D
Users of typical antipsychotics
Users of atypical antipsychotics
n per group
ASAMD
p-value
Typical antipsychotics better
Atypical anti-psychotics better
Risk set sampling and propensity score matching 13 26 186 0.053 0.03
Risk set sampling only 19 32 265 0.082 0.06
Number of patients requiring mechanical ventilation within 30 days
Odds ratio (95% CI)

Acknowledgments:

The authors acknowledge their partners and families for support in childcare and other matters during this time. The views, opinions, and findings contained in this study are those of the authors and do not represent the official views, policies, or endorsement of the Department of Defense or the U.S. Government. Imaging at the University of Sheffield was performed in the Wolfson Light Microscopy facility. A549 cells stably expressing ACE2 (A549-ACE2) were provided by O. Schwartz. We thank H. Malik for helpful discussions. We thank M. Akdel and J. Durairaj for input on analysis, R. Albrecht for support with the BSL3 facility and procedures at the ISMMS, and R. Cadagan for technical assistance. We thank the High Containment Core at Georgia State University for supporting the BSL3 facility and procedures. We thank B. L. Roth, A. Alon, and A. C. Kruse for helpful discussions on sigma receptor biology. We thank S. Deeks, P. J. Hunt, J. Gordan, C. Corbetta-Rastelli, and E. Lantos for insight on clinical applications for drug repurposing. Funding: This research was funded by grants from the National Institutes of Health (NIH) (P01AI063302, P50AI150476, R01AI120694, R01AI122747, R01AI143292, U19AI135972, and U19AI135990 to N.J.K.; P01AI120943 and R01AI143292 to C.F.B.; and U19 AI135990 to T.I.); by the National Institute of Allergy and Infectious Diseases (NIAID) R01AI128214 to O.S.R.; by the National Institute of Neurological Disorders and Stroke R01 NS089713, the NIH Office of the Director AI150476, and NIGMS R01 GM24485 to R.M.S.; by a Fast Grant for COVID-19 from the Emergent Ventures program at the Mercatus Center of George Mason University (N.J.K.) and a separate Fast Grant for COVID-19 (C.F.B.); by the Roddenberry Foundation Gladstone Institutes to K.S.P. and N.J.K.; from the Defense Advanced Research Projects Agency (HR0011-19-2-0020 to B.S., N.J.K., K.A.V., D.A.A., A.G.-S., and K.M.S.; HR0011-20-2-0040 to M.B.F.); NIGMS R35GM122481 (to B.S.); by CRIP (Center for Research for Influenza Pathogenesis), a NIAID supported Center of Excellence for Influenza Research and Surveillance (CEIRS, contract no. HHSN272201400008C) to A.G.-S.; by supplements to NIAID grant U19AI135972 and DoD grant W81XWH-20-1-0270 to A.G.-S.; by the Bill and Melinda Gates Foundation (INV-006099) and BARDA (ASPR-20-01495) to M.B.F.; by Howard Hughes Medical Institute to K.S.; by Damon Runyon Cancer Research Foundation DRG-2402-20 to C.P.; by Burroughs Wellcome Fund 1019894 to N.He.; by the Chan Zuckerberg Initiative to O.S.R. and T.Ko.; by Cytoscape: A modeling platform for biomolecular networks (NHGRI R01 HG009979) to T.I.; by the generous support of the JPB Foundation, the Open Philanthropy Project [research grant 2020-215611 (5384)] and by anonymous donors to A.G.-S.; a Laboratoire d'Excellence grant ANR-10-LABX-62-IBEID and the URGENCE COVID-19 Institut Pasteur fundraising campaign to M.V and N.Jo.; by grants from the BBSRC (BB/S009566/1 and BB/L002841/1) to A.A.P. and D.W. and by BBSRC

White Rose DTP (BB/J014443/1) to A.S.S.-S.; by The Augusta University–Georgia State University Seed Grant program to C.F.B.; by the MRC grant MC PC 19026 to M.P.; by the MRC Grant MC_UU_12016/2 to D.A.; by the Medical Research Council (MC_UU_12016/2) to D.R.A.; by the DFG under Germany's Excellence Strategy (EXC-2189, project ID 390939984 to R.G.); and funding from F. Hoffmann-La Roche and Vir Biotechnology and gifts from The Ron Conway Family and Vir Biotechnology to the Quantitative Biosciences Institute Coronavirus Research Group (QCRG). This work was supported by the Excellence in Research Award (ERA) from the Laboratory for Genomics Research, a collaboration between the Innovative Genomics Institute of UC Berkeley/UCSF and GlaxoSmithKline (133122P to M.O. and N.J.K.). J.H. was supported by the UCSF Medical Scientist Training Program (T32GM007618); P.F. was supported by the UCSF Medical Scientist Training Program (T32GM007618) and the NIH-NIAID (F30AI143401); M.B. was supported by the NCI at the NIH (F32CA239333); H.T.K. was supported by the NIH (K99GM138753); K.H. was supported by the National Science Foundation (1650113); B.T. was supported by the NIH (F32 CA239336); U.S.C. was supported by the National Institute of General Medical Sciences (F32GM137463). The QCRG Structural Biology Consortium has received support from: Quantitative Biosciences Institute, Defense Advance Research Projects Agency [HR0011-19-2-0020 to D.A.A. and K.A.V.; B.S., primary investigator (PI)], FastGrants COVID-19 grant (K.A.V., PI), Laboratory For Genomics Research (O.S.R., PI), and Laboratory For Genomics Research (R.M.S., PI). Author contributions: The following authors designed and conceptualized the study: A.A.P., A.G.-S., C.F.B., D.E.G., H.B., J.A.R., J.Ba., J.H., K.A.V., K.O., M.B., M.V., N.J.K., O.S.R., P.B., R.G., and R.M.K. The following authors performed experiments or data acquisition: A.A.P., A.Du., A.G., A.J.H., A.L.L., A.R., A.R.W., A.S.J., A.S.S.-S., A.V.-G., B.T., C.J.H., C.A., C.Ba., C.F.B., C.J., C.K., C.R.S., D.B., D.E.G., D.K., D.L.S., D.M., D.M.W., D.R.A., D.S., E.M., E.P., E.H.P., E.R., E.W.T., F.Be., F.Br., G.J., G.K., H.F., H.L., I.B.-H., I.Du., J.A.R., J.A.S., J.Ba., J.C.-S., J.H., J.Lo., J.O., J.X., J.Z.G., K.A.B., K.C.K., K.H., K.M.W., L.M., M.A.K., M.B., M.B.F., M.Ch., M.C.M., M.D., M.El., M.Gu., M.Mod., M.M.K., M.McGra., M.McGre., M.U., M.V., N.G., N.Jo., P.B., P.D., P.F., R.Hü., R.Ha., R.J.K., R.M.K., R.R., R.To., S.G.W., S.P., S.Ra., S.U., S.Wei., S.Wes., T.Ke., T.M., T.V., T.W.O., V.V.R., Y.S., and Z.Z. The following authors conducted formal data analysis: A.A.P., A.Du., A.J.H., A.L.L., A.R.W., A.S., A.S.J., A.S.S.-S., A.V.-G., B.S., C.J.H., C.A., C.F.B., C.K., D.B., D.E.G., D.L.S., D.M., D.M.W., D.S., E.M., E.P., E.W.T., G.J., G.K., H.B., I.Du., J.A.R., J.A.S., J.Ba., J.H., J.Par., K.A.V., K.C.K., K.M.W., K.O., K.R.H., K.S.P., L.M., M.A.K., M.B., M.B.F., M.Ca., M.C.M., M.Ec., M.Gh., M.Gu., M.J.O., M.M.K., M.McGra., M.V., N.Jo., N.Ju., O.S.R., P.B., R.G., R.Hü., R.M.K., R.R., S.P., S.U., S.Wei., T.A.T., T.I., T.V., T.W.O., U.R., V.V.R., Y.Zho., Z.C., and Z.Z. The following authors supervised or

managed research: A.A.P., A.F., A.G.-S., A.M., A.S., B.S., C.A., C.F.B., D.A.A., D.E.G., D.L.S., J.A.R., J.H., J.M.F., J.S.F., K.A.V., K.H., K.M.S., K.M.W., K.O., K.S.P., M.B., M.B.F., M.Ch., M.O., M.P., M.V., N.Jo., N.Ju., N.J.K., O.S.R., P.B., R.G., R.Hü., R.M.K., T.I., T.Ko., and Y.C. The following authors raised funds for these efforts: A.A.P., A.G.-S., B.S., C.F.B., D.A.A., D.E.G., D.R.A., J.M.F., K.M.S., K.O., K.S.P., M.P., M.V., N.Jo., N.J.K., O.S.R., P.B., R.G., and T.I. The following authors drafted the original manuscript: A.A.P., B.S., C.A., D.E.G., D.L.S., H.B., J.A.R., J.Ba., J.H., K.A.V., K.M.H., K.M.S., K.O., K.R.H., K.S.P., M.A.K., M.B., M.Ca., M.Ec., M.M.K., M.O., M.P., M.So., N.J.K., O.S.R., P.B., P.H., R.M.K., T.A.T., U.R., V.V.R., and Z.Z.C.N. The structural biology portion of this work was performed by the Quantitative Biosciences Institute Coronavirus Research Group Structural Biology Consortium. Listed below are the contributing members of the consortium, listed by teams in order of team relevance to the published work. The team leads are listed first (those responsible for organization and the experimental design utilized within each team), and the rest of team members are listed alphabetically. Bacterial expression team: A.D., M.G., E.W.T., J.C., L.D., S.F., M.J., H.T.K., V.L.L., Y.L., M.L., G.E.M., J.P., A.C.T., Z.Y., F.Z., and Y.Z. Protein purification team: M.M., T.W.O., S.P., C.M.A., C.M.C., B.F., M.G., K.K., J.P., J.K.P., K.S., and T.K.M.T. Cryo-EM grid freezing and collection team: C.M.A., A.F.B., G.E.M., C.P., A.N.R., M.S., J.R.B., M.G., F.L., K.E.L., A.M., F.M., J.P., T.H.P., Jr., S.P., A.M.S., P.V.T., F.W., and Z.Y. Cryo-EM data processing team: A.F.B., M.S.D., G.E.M., H.C.N., A.N.R., D.A., J.R.B., M.G.C., C.M.C., U.S.C., D.D., B.F., M.G., N.H., M.J., F.L., J.L., Y.L., J.P., T.H.P., Jr., S.P., S.S., R.T., D.T., E.T., K.Z., and F.Z. Mammalian cell expression team: C.B., M.G.C., D.D., C.N., A.M.S., J.Z., C.M.A., A.B., N.H., Y.L., P.N., C.P., M.S., S.S., K.S., R.T., T.K.M.T., and N.W. Crystallography team: N.H., H.T.K., U.S.-G., I.D.Y., J.B., I.D., and X.L. Infrastructure team: D.B., A.J., A.J., L.M., M.T., and E.T. Leadership team: O.S.R., K.A.V., D.A.A., Y.C., J.S.F., A.F., N.J., T.K., N.J.K., A.M., D.R.S., and R.M.S. All authors edited the manuscript. Competing interests: The Krogan Laboratory has received research support from Vir Biotechnology and F. Hoffmann-La Roche. T.I. is the cofounder of Data4Cure, Inc., with an equity interest, and he has a funded sponsored research agreement from Ideaya BioSciences, Inc., with an equity stake. The García-Sastre Laboratory has received research support from Pfizer, Senhwa Biosciences, and 7Hills Pharma; and A.G.-S. has consulting agreements for the following companies involving cash and/or stock: Vivaldi Biosciences, Contrafect, 7Hills Pharma, Avimex, Valneva, Accurius, and Esperovax. K.M.S. has consulting agreements for the following companies involving cash and/or stock compensation: Black Diamond Therapeutics, BridGene Biosciences, Denali Therapeutics, Dice Molecules, eFFECTOR Therapeutics, Erasca, Genentech/Roche, Janssen Pharmaceuticals, Kumquat Biosciences, Kura Oncology, Merck,

Mitokinin, Petra Pharma, Qulab Inc. Revolution Medicines, Type6 Therapeutics, Venthera, and Wellspring Biosciences (Araxes Pharma). J.A.R. is an employee and shareholder of Aetion, Inc., a company that makes software for the analysis of real-world data. A.R.W. is an employee and shareholder of Aetion, Inc., a company that makes software for the analysis of real-world data. R.J.K. is an employee of HealthVerity, a company that links and deidentifies real-world data. J.S.F. is a founder of Keyhole Therapeutics and a shareholder of Relay Therapeutics and Keyhole Therapeutics. The Fraser laboratory has received sponsored research support from Relay Therapeutics. K.H., J.C.-S., J.O., and T.M. are employees and shareholders of Synthego Corporation. Aetion holds patents related to this work, including U.S. Patent no. 9,378,271; other patents are pending. Data and materials availability: Further information and requests for resources and reagents should be directed to and will be fulfilled by N.J.K. (nevan.krogan@ucsf.edu). The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD021588 (100). An interactive version of PPI data can be found at https://kroganlab.ucsf.edu/network-maps. Atomic coordinates and the cryo-EM map of the reported Tom70-ORF9b structure have been deposited in the Protein Data Bank under accession code 7KDT and in the Electron Microscopy Data Bank under accession code EMD-22829. Expression vectors used in this study are readily available from the authors for biomedical researchers and educators in the nonprofit sector. The Aetion Evidence Platform used for the clinical analysis is available under license from Aetion, New York, NY. To protect patient privacy, data used in real-world analyses are available for inspection by qualified researchers under confidentiality and third-party agreements with Aetion and/or HealthVerity.

Supplementary Materials

Quantitative Biosciences Institute Coronavirus Research Group, Structural Biology Consortium: Tristan W. Owens1, Meghna Gupta1, Sergei Pourmal1, Erron W. Titus1, Caleigh M. Azumaya1, Christian Billesbølle1, Axel F. Brilot1, Melody G. Campbell1,2, Amy Diallo1, Miles Sasha Dickinson1, Devan Diwanji1, Nadia Herrera1, Nick Hoppe1, Huong T. Kratochvil1, Yanxin Liu1, Gregory E. Merz1, Michelle Moritz1, Henry C. Nguyen1, Carlos Nowotny1, Cristina Puchades1, Alexandrea N. Rizo1, Ursula Schulze-Gahmen1, Amber M. Smith1, Ming Sun1,3, Iris D. Young1, Jianhua Zhao1, Daniel Asarnow1, Justin Biel1, Alisa Bowen1, Julian R. Braxton1, Jen Chen1, Cynthia M. Chio1, Un Seng Chio1, Ishan Deshpande1, Loan Doan1, Bryan Faust1, Sebastian Flores1, Mingliang Jin1, Kate Kim1, Victor L. Lam1, Fei Li1, Junrui Li1, Yen-Li Li1, Yang Li1, Xi Liu1, Megan Lo1, Kyle E. Lopez1, Arthur A. Melo1, Frank R. Moss III1, Phuong Nguyen1, Joana Paulino1, Komal Ishwar Pawar1, Jessica K. Peters1, Thomas H. Pospiech Jr.1, Maliheh Safari1, Smriti Sangwan1, Kaitlin Schaefer1, Paul V. Thomas1, Aye C. Thwin1, Raphael Trenker1, Eric Tse1, Tsz Kin Martin Tsui1, Feng Wang1, Natalie Whitis1, Zanlin Yu1, Kaihua Zhang1, Yang Zhang1, Fengbo Zhou1, Donovan Trinidad1, David A. Agard1,4–6, Yifan Cheng1,4–7, James S. Fraser1,4,5,8, Adam Frost1,4–6, Natalia Jura1,4,5,9,10, Tanja Kortemme1,4,5,8,11, Aashish Manglik1,4,5,12, Daniel R. Southworth1,5,6, Robert M. Stroud1,4–6, Oren S. Rosenberg1,4–6,13,14, Kliment A. Verba1,4,5,12

Affiliations:

1. Quantitative Biosciences Institute (QBI) Coronavirus Research Group Structural Biology Consortium, University of California, San Francisco, CA, 94158, USA.

2. Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.

3. Beam Therapeutics, Cambridge, MA, 02139, USA.

4. Quantitative Biosciences Institute (QBI), University of California, San Francisco, CA, 94158, USA.

5. QBI COVID-19 Research Group (QCRG), San Francisco, CA 94158, USA.

6. Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, CA, 94158, USA.

7. Howard Hughes Medical Institute.

8. Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, 94158, USA.

9. Cardiovascular Research Institute, University of California San Francisco, San Francisco, CA, 94158, USA.

10. Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA, 94158, USA.

11. The UC Berkeley-UCSF Graduate Program in Bioengineering, University of California, San Francisco, CA, 94158, USA.

12. Department of Pharmaceutical Chemistry, University of California, San Francisco, 94158, USA.

13. Department of Medicine, University of California, San Francisco, CA, 94143 USA.

14. Chan-Zuckerberg Biohub, San Francisco, CA, 94158, USA.

Zoonomia Consortium

Joana Damas1, Graham M. Hughes2, Kathleen C. Keough3,4, Corrie A. Painter5, Nicole S. Persky6, Marco Corbo1, Bodgan Kirilenko7-9, Michael Hiller7-9, Klaus-Peter Koepfli10, Irene Kaplow11, Morgan Wirthlin11, Andreas R. Pfenning11, Huabin Zhao12, 13 Diane P. Genereux14, Ross Swofford14, Abigail Lind4,15,16, Katherine S. Pollard4,15,16, Oliver A. Ryderq17,18, Martin T. Nweeia19,20,21, Jennifer Meadows14,22, Michael Dong14,22, Ola Wallerman14,22, Vikki Marinescu14,22, Kerstin Lindblad-Toh14,22, David A. Ray23, Sarahjane Power2, Emma C. Teeling2, Gaurav Chauhan14,24,25, Shirley Xue Li14,24,25, Elinor K. Karlsson14,24,25, and Harris A. Lewin1,26,27

Affiliations:

1. The Genome Center, University of California, Davis, CA 95616

2. School of Biology and Environmental Science, University College Dublin, Belfield, Dublin 4, Ireland

3. Graduate Program in Pharmaceutical Sciences and Pharmacogenomics, Quantitative Biosciences Consortium, University of California, San Francisco, CA 94117

4. Gladstone Institute of Data Science and Biotechnology, San Francisco, CA 94158

5. Cancer Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142

6. Genetic Perturbation Platform, Broad Institute of MIT and Harvard, Cambridge, MA 02142

7. Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany

8. Max Planck Institute for the Physics of Complex Systems, 01187 Dresden, Germany

9. Center for Systems Biology Dresden, 01307 Dresden, Germany

10. Center for Species Survival, Smithsonian Conservation Biology Institute, National Zoological Park, Front Royal, VA 22630

11. Department of Computational Biology, School of Computer Science,Carnegie Mellon University, Pittsburgh, PA 15213

12. Department of Ecology, Tibetan Centre for Ecology and Conservation at WHU-TU, Hubei Key Laboratory of Cell Homeostasis, College of Life Sciences, Wuhan University, Wuhan 430072, China

13. College of Science, Tibet University, Lhasa 850000, China

14. Broad Institute of MIT and Harvard, Cambridge, MA 02142

15. Department of Epidemiology & Biostatistics, Institute for Computational Health Sciences, and Institute for Human Genetics, University of California, San Francisco, CA 94158

16. Chan Zuckerberg Biohub, San Francisco, CA 94158

17. San Diego Zoo Institute for Conservation Research, Escondido, CA 92027

18. Department of Evolution, Behavior, and Ecology, Division of Biology, University of California San Diego, La Jolla, CA 92093

19. Department of Restorative Dentistry and Biomaterials Sciences, Harvard School of Dental Medicine, Boston, MA 02115

20. School of Dental Medicine, Case Western Reserve University, Cleveland, OH 44106

21. Marine Mammal Program, Department of Vertebrate Zoology, Smithsonian Institution, Washington, DC 20002

22. Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, 751 23 Uppsala, Sweden

23. Department of Biological Sciences, Texas Tech University, Lubbock, TX, 79409

24. Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01655

25. Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA 01655

26. Department of Evolution and Ecology, University of California, Davis, CA 95616

27. John Muir Institute for the Environment, University of California, Davis, CA 95616

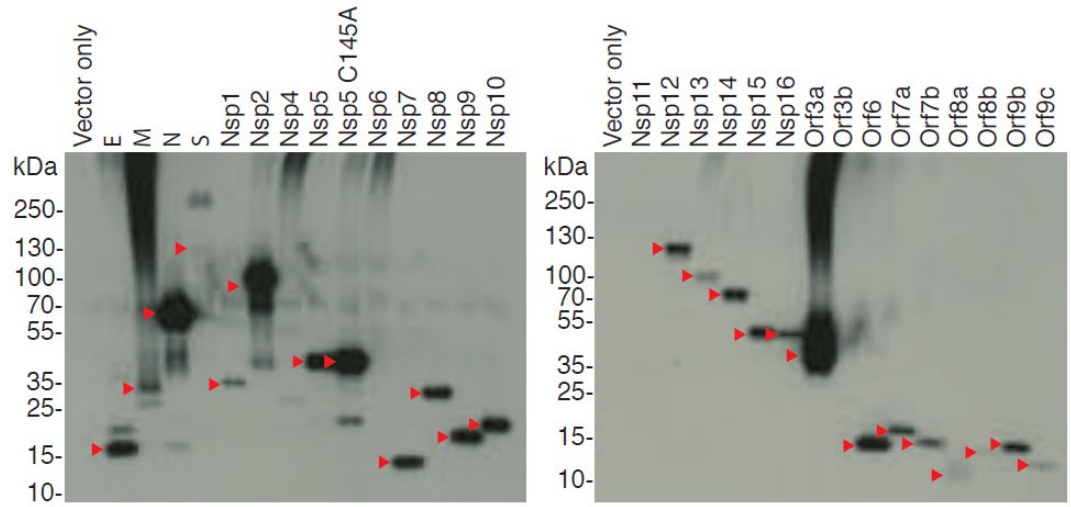


Figure S1. SARS-CoV-1 protein expression.

Input samples from immunoprecipitations were probed by immunoblot using anti-Strep antibody. Red arrowhead indicates that the band appears near expected molecular weight. Nsp = non-structural protein; Orf = open reading frame.

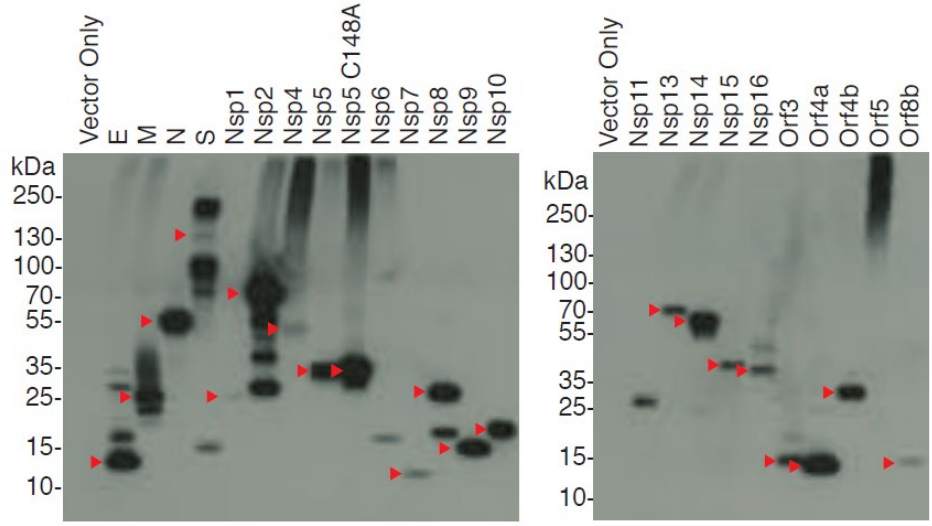


Figure S2. MERS-CoV protein expression.

Input samples from immunoprecipitations were probed by immunoblot using anti-Strep antibody. Red arrowhead indicates that the band appears near expected molecular weight. Nsp = non-structural protein; Orf = open reading frame.

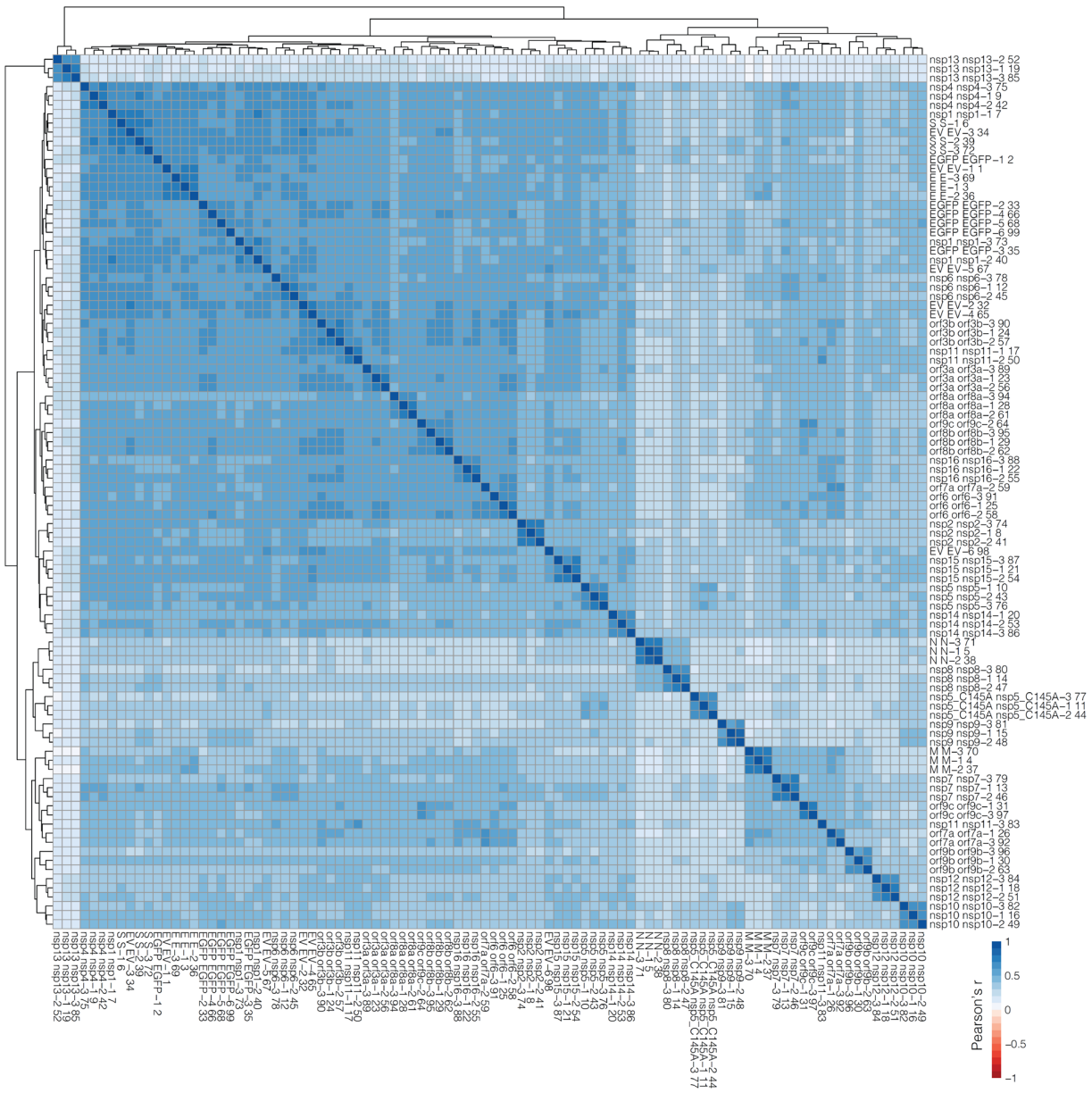


Figure S3. Correlation analysis of SARS-CoV-1 proteomics samples.

Pearson's pairwise correlations were calculated for all combinations of replicates of SARS-CoV-1 affinity purification-mass spectrometry (AP-MS) samples. Unbiased clustering was applied and correlation scores are depicted by heatmap. All MS samples were compared and clustered using standard artMS (https://github.com/biodavidjm/artMS) procedures on observed feature intensities computed by MaxQuant.

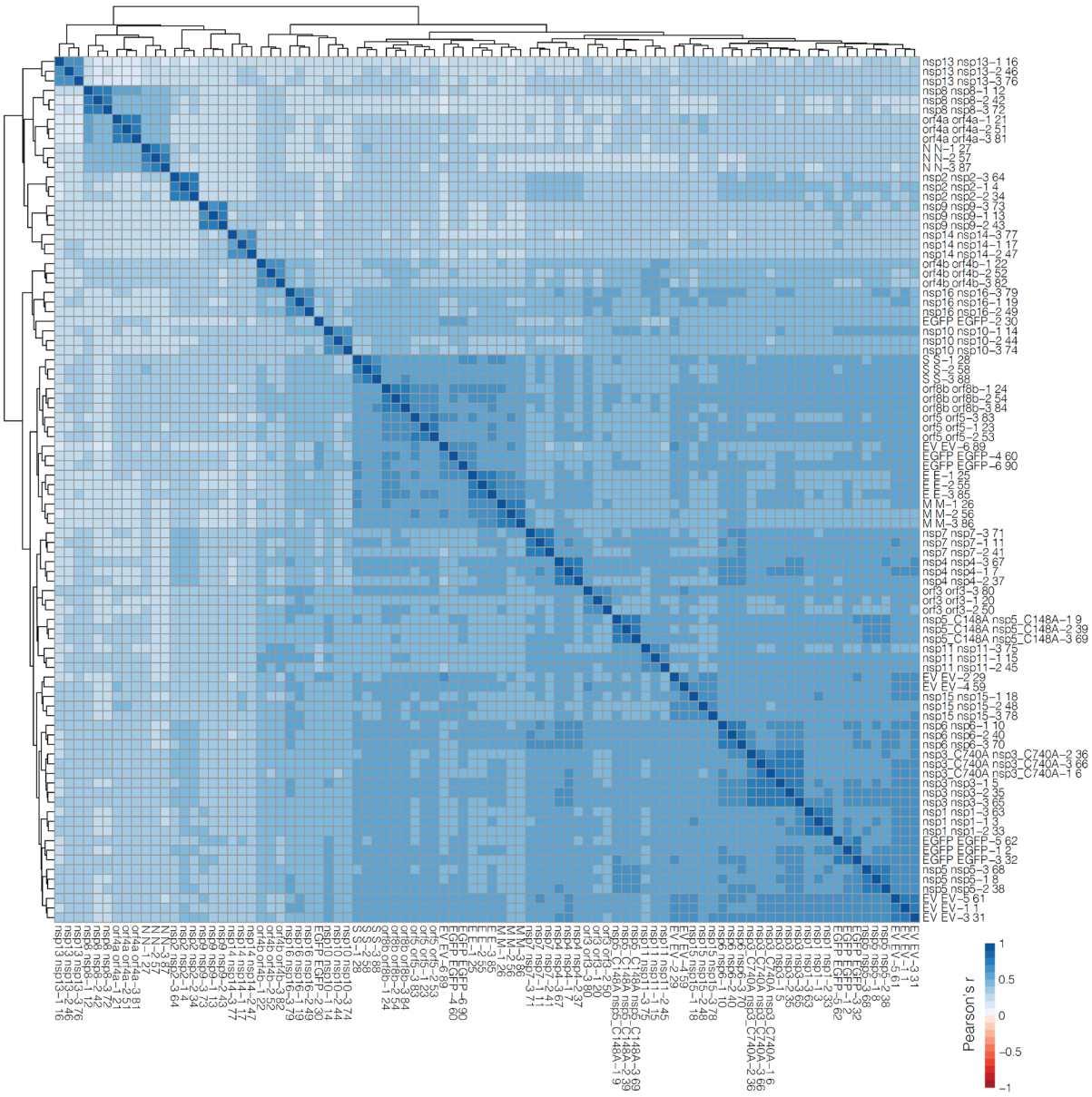


Figure S4. Correlation analysis of MERS-CoV proteomics samples. Pearson's pairwise correlations were calculated for all combinations of replicates of MERS-CoV affinity purification-mass spectrometry (AP-MS) samples. Unbiased clustering was applied and correlation scores are depicted by heatmap. All MS samples were compared and clustered using standard artMS (https://github.com/biodavidjm/artMS) procedures on observed feature intensities computed by MaxQuant.

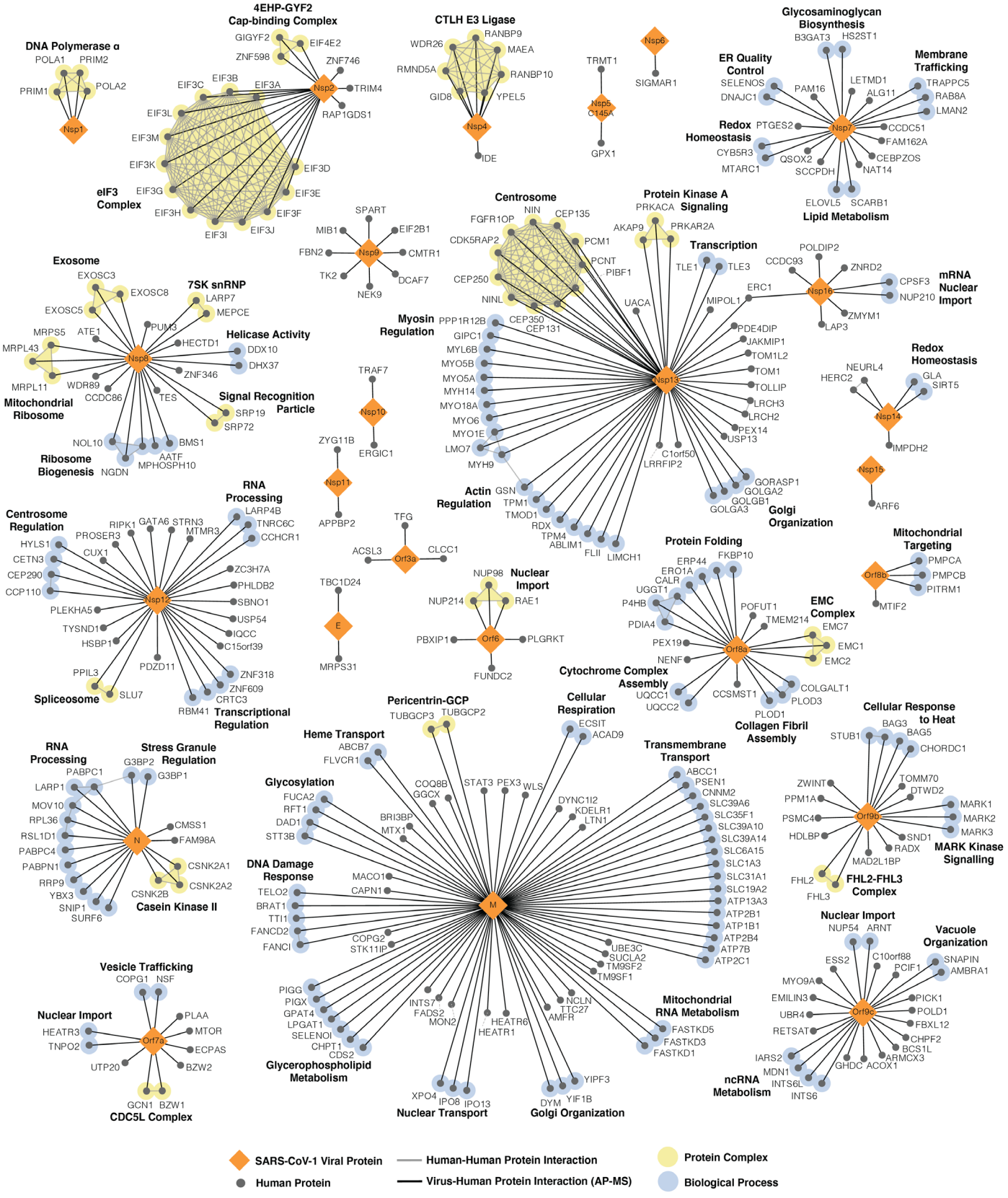

DNA Polymerase α
POLA1
PRIM2
PRIM1
POLA2
Nsp1
4EHP-GYF2
Cap-binding Complex
GIGYF2
EIF4E2
ZNF598
ZNF746
EIF3B
EIF3C
EIF3A
Nsp2
TRIM4
RAP1GDS1
EIF3L
EIF3M
EIF3K
EIF3D
eIF3
Complex
EIF3G
EIF3E
EIF3H
EIF3F
EIF3I
EIF3J
CTLH E3 Ligase
RANBP9
WDR26
MAEA
RMND5A
RANBP10
GID8
YPEL5
Nsp4
IDE
TRMT1
Nsp5
C145A
GPX1
Nsp6
SIGMAR1
Glycosaminoglycan
Biosynthesis
B3GAT3
HS2ST1
ER Quality
Control
SELENOS
DNAJC1
PAM16
LETMD1
ALG11
Membrane
Trafficking
TRAPPC5
RAB8A
LMAN2
Redox
Homeostasis
PTGES2
Nsp7
CCDC51
FAM162A
CYB5R3
QSOX2
CEBPZOS
MTARC1
SCCPDH
NAT14
ELOVL5
SCARB1
Lipid Metabolism
Centrosome
NIN
CEP135
FGFR1OP
PCM1
CDK5RAP2
PCNT
PIBF1
CEP250
NINL
Protein Kinase A
Signaling
PRKACA
PRKAR2A
AKAP9
Transcription
TLE1
TLE3
POLDIP2
CCDC93
ZNRD2
ERC1
Nsp16
CPSF3
NUP210
MIPOL1
ZMYM1
LAP3
mRNA
Nuclear
Import
UACA
Myosin
Regulation
PPP1R12B
CEP350
CEP131
GIPC1
MYL6B
MYO5B
MYO5A
MYH14
MYO18A
MYO6
MYO1E
LMO7
MYH9
Nsp13
PDE4DIP
JAKMIP1
TOM1L2
TOM1
TOLLIP
LRCH3
LRCH2
PEX14
USP13
C1orf50
LRRFIP2
Exosome
EXOSC3
EXOSC8
EXOSC5
7SK snRNP
LARP7
MEPCE
PUM3
MRPS5
ATE1
MRPL43
HECTD1
Helicase Activity
DDX10
DHX37
Nsp8
ZNF346
MRPL11
WDR89
CCDC86
TES
Mitochondrial
Ribosome
Signal Recognition
Particle
SRP19
SRP72
NOL10
BMS1
AATF
MPHOSPH10
NGDN
Ribosome
Biogenesis
SPART
MIB1
EIF2B1
FBN2
Nsp9
CMTR1
TK2
DCAF7
NEK9
TRAF7
Nsp10
ZYG11B
ERGIC1
Nsp11
APPBP2
Redox
Homeostasis
NEURL4
HERC2
GLA
SIRT5
Nsp14
IMPDH2
Nsp15
ARF6
Actin
Regulation
GSN
TPM1
TMOD1
RDX
TPM4
ABLIM1
FLII
LIMCH1
GORASP1
GOLGA2
GOLGB1
GOLGA3
Golgi
Organization
RNA
Processing
LARP4B
TNRC6C
CCHCR1
Centrosome
Regulation
RIPK1
GATA6
STRN3
MTMR3
PROSER3
HYLS1
CUX1
CETN3
CEP290
CCP110
Nsp12
ZC3H7A
PHLDB2
SBNO1
USP54
IQCC
C15orf39
PLEKHA5
TYSND1
HSBP1
PDZD11
PPIL3
SLU7
Spliceosome
ZNF318
ZNF609
CRTC3
RBM41
Transcriptional
Regulation
TFG
ACSL3
Orf3a
CLCC1
TBC1D24
E
MRPS31
NUP98
Nuclear
Import
NUP214
RAE1
PBXIP1
Orf6
PLGRKT
FUNDC2
Protein Folding
ERP44
FKBP10
ERO1A
CALR
UGGT1
P4HB
PDIA4
POFUT1
TMEM214
EMC
Complex
EMC7
EMC1
EMC2
PEX19
Orf8a
NENF
Cytochrome Complex
Assembly
UQCC1
UQCC2
CCSMST1
COLGALT1
PLOD3
PLOD1
Collagen Fibril
Assembly
Mitochondrial
Targeting
Orf8b
PMPCA
PMPCB
PITRM1
MTIF2
RNA
Processing
Stress Granule
Regulation
G3BP2
G3BP1
PABPC1
LARP1
MOV10
RPL36
RSL1D1
PABPC4
PABPN1
RRP9
YBX3
SNIP1
SURF6
N
CMSS1
FAM98A
CSNK2A1
CSNK2A2
CSNK2B
Casein Kinase II
Pericentrin-GCP
TUBGCP3
TUBGCP2
Heme Transport
ABCB7
FLVCR1
Cellular
Respiration
ECSIT
ACAD9
Transmembrane
Transport
ABCC1
PSEN1
CNNM2
SLC39A6
SLC35F1
SLC39A10
SLC39A14
SLC6A15
SLC1A3
SLC31A1
SLC19A2
ATP13A3
ATP2B1
ATP1B1
ATP2B4
ATP7B
ATP2C1
Glycosylation
FUCA2
RFT1
DAD1
STT3B
COQ8B
GGCX
STAT3
PEX3
WLS
DYNC1I2
KDELR1
LTN1
BRI3BP
MTX1
DNA Damage
Response
MACO1
CAPN1
TELO2
BRAT1
TTI1
FANCD2
FANCI
M
COPG2
STK11IP
UBE3C
SUCLA2
TM9SF2
TM9SF1
Cellular Response
to Heat
BAG3
BAG5
STUB1
CHORDC1
ZWINT
TOMM70
DTWD2
PPM1A
Orf9b
MARK1
MARK2
MARK3
PSMC4
HDLBP
SND1
RADX
MAD2L1BP
MARK Kinase
Signalling
FHL2
FHL3
FHL2-FHL3
Complex
Vesicle Trafficking
COPG1
NSF
Nuclear Import
HEATR3
TNPO2
Orf7a
PLAA
MTOR
ECPAS
BZW2
UTP20
GCN1
BZW1
CDC5L Complex
PIGG
PIGX
GPAT4
LPGAT1
SELENOI
CHPT1
CDS2
Glycerophospholipid
Metabolism
INTS7
FADS2
MON2
HEATR6
HEATR1
NCLN
TTC27
AMFR
Mitochondrial
RNA Metabolism
FASTKD5
FASTKD3
FASTKD1
XPO4
IPO8
IPO13
Nuclear Transport
YIPF3
DYM
YIF1B
Golgi Organization
Nuclear Import
NUP54
ARNT
Vacuole
Organization
ESS2
C10orf88
PCIF1
SNAPIN
AMBRA1
MYO9A
Orf9c
EMILIN3
PICK1
UBR4
POLD1
RETSAT
FBXL12
CHPF2
BCS1L
IARS2
ARMCX3
GHDC
ACOX1
MDN1
INTS6L
INTS6
ncRNA
Metabolism
SARS-CoV-1 Viral Protein
Human Protein
Human-Human Protein Interaction
Virus-Human Protein Interaction (AP-MS)
Protein Complex
Biological Process

Figure S5. SARS-CoV-1 Virus-Human Protein Interaction Network.

Virus-human protein- protein interaction map depicting high-confidence interactions (MiST ≥ 0.7 & Saint BFDR ≤0.05 & Average Spectral Counts ≥ 2) for SARS-CoV-1 as derived from affinity purification- mass spectrometry (AP-MS). Viral bait proteins are depicted with orange diamonds and human proteins with dark grey circles. Human-human interactions are depicted in thin, dark grey lines. Proteins within the same protein complexes or biological process are indicated with light yellow or light blue highlighting, respectively, and annotated accordingly. Human-human physical interactions, protein complex definitions, and biological process groupings are derived from CORUM (46), Gene Ontology (biological process), and manually curated from literature sources. Thin dashed grey lines are used to indicate the placement of node labels when adjacent node labels would have otherwise been obscured.

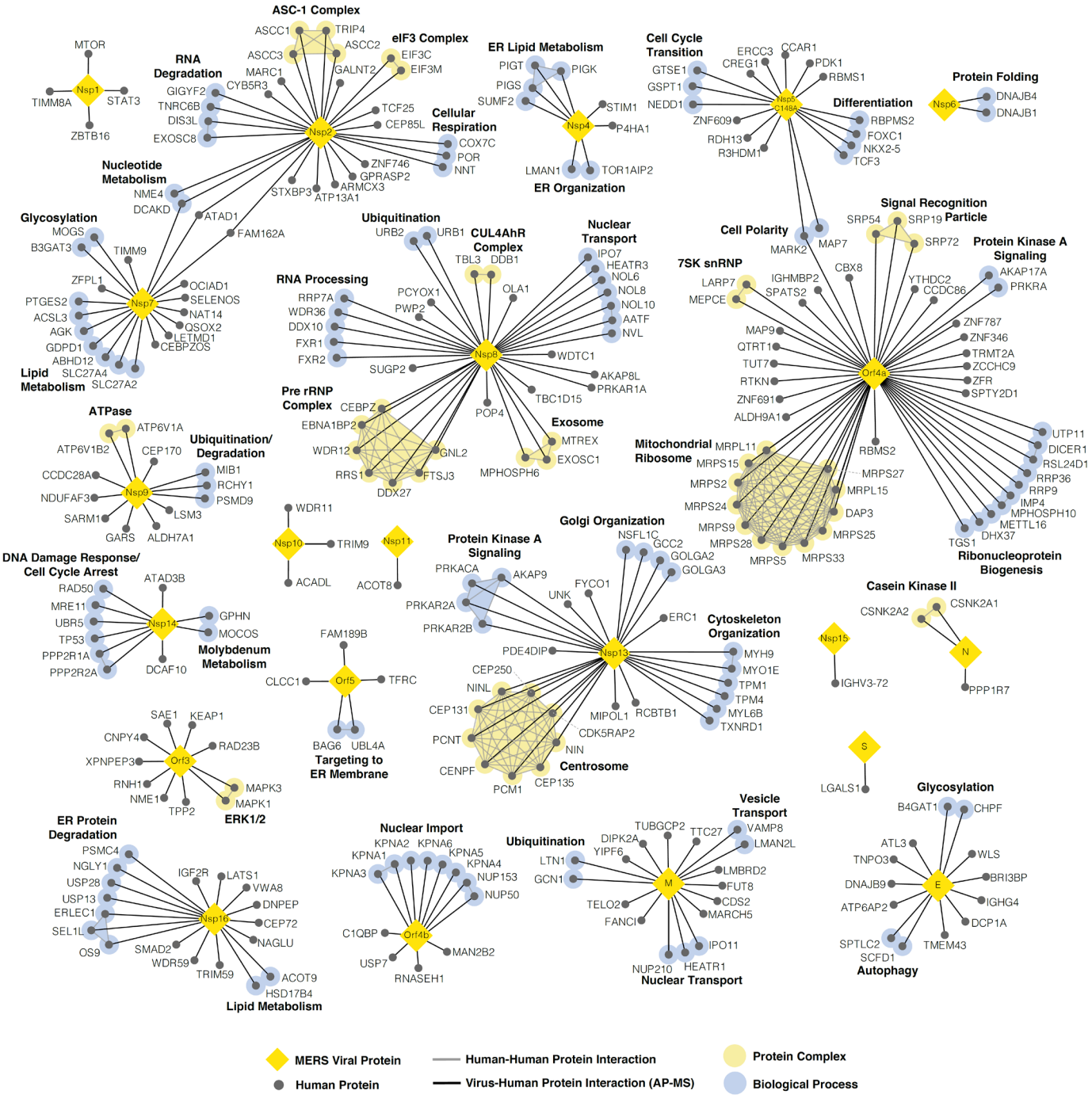

ASC-1 Complex
ASCC1
TRIP4
ASCC2
ASCC3
eIF3 Complex
EIF3C
EIF3M
MTOR
Nsp1
TIMM8A
STAT3
ZBTB16
RNA Degradation
MARC1
CYB5R3
GIGYF2
TNRC6B
DIS3L
EXOSC8
GALNT2
Nsp2
TCF25
CEP85L
Cellular Respiration
COX7C
POR
NNT
ZNF746
GPRASP2
ARMCX3
ATP13A1
STXBP3
ATAD1
FAM162A
Nucleotide Metabolism
NME4
DCAKD
ER Lipid Metabolism
PIGT
PIGK
PIGS
SUMF2
Nsp4
STIM1
P4HA1
LMAN1
TOR1AIP2
ER Organization
Cell Cycle Transition
GTSE1
GSPT1
NEDD1
ERCC3
CCAR1
CREG1
PDK1
RBMS1
Nsp5 C148A
ZNF609
RDH13
R3HDM1
Differentiation
RBPMS2
FOXC1
NKX2-5
TCF3
Protein Folding
Nsp6
DNAJB4
DNAJB1
Glycosylation
MOGS
B3GAT3
TIMM9
ZFPL1
Nsp7
OCIAD1
SELENOS
NAT14
QSOX2
LETMD1
CEBPZOS
PTGES2
ACSL3
AGK
GDPD1
ABHD12
SLC27A4
SLC27A2
Lipid Metabolism
Ubiquitination
URB2
URB1
CUL4AhR Complex
TBL3
DDB1
OLA1
Nuclear Transport
IPO7
HEATR3
NOL6
NOL8
NOL10
AATF
NVL
RNA Processing
RRP7A
WDR36
DDX10
FXR1
FXR2
PCYOX1
PWP2
Nsp8
WDTC1
SUGP2
AKAP8L
PRKAR1A
TBC1D15
POP4
Pre rRNP Complex
CEBPZ
EBNA1BP2
WDR12
GNL2
RRS1
FTSJ3
DDX27
Exosome
MTREX
EXOSC1
MPHOSPH6
Cell Polarity
MARK2
MAP7
Signal Recognition Particle
SRP54
SRP19
SRP72
7SK snRNP
LARP7
MEPCE
IGHMBP2
SPATS2
CBX8
Protein Kinase A Signaling
AKAP17A
PRKRA
YTHDC2
CCDC86
Orf4a
MAP9
QTRT1
TUT7
RTKN
ZNF691
ALDH9A1
ZNF787
ZNF346
TRMT2A
ZCCHC9
ZFR
SPTY2D1
ATPase
ATP6V1A
ATP6V1B2
CEP170
Ubiquitination/ Degradation
MIB1
RCHY1
PSMD9
LSM3
ALDH7A1
GARS
SARM1
NDUFAF3
CCDC28A
Nsp9
Mitochondrial Ribosome
MRPL11
MRPS15
MRPS2
MRPS24
MRPS9
MRPS28
MRPS5
MRPS33
MRPS25
DAP3
MRPL15
MRPS27
RBMS2
UTP11
DICER1
RSL24D1
RRP36
RRP9
IMP4
MPHOSPH10
METTL16
DHX37
TGS1
Ribonucleoprotein Biogenesis
WDR11
Nsp10
TRIM9
ACADL
Nsp11
ACOT8
Golgi Organization
NSFL1C
GCC2
GOLGA2
GOLGA3
Protein Kinase A Signaling
PRKACA
AKAP9
PRKAR2A
PRKAR2B
UNK
FYCO1
ERC1
Cytoskeleton Organization
PDE4DIP
Nsp13
MYH9
MYO1E
TPM1
TPM4
MYL6B
TXNRD1
CEP250
NINL
CEP131
PCNT
CENPF
PCM1
CEP135
MIPOL1
RCBTB1
CDK5RAP2
NIN
Centrosome
DNA Damage Response/ Cell Cycle Arrest
RAD50
MRE11
UBR5
TP53
PPP2R1A
PPP2R2A
ATAD3B
Nsp14
GPHN
MOCOS
DCAF10
Molybdenum Metabolism
FAM189B
Orf5
CLCC1
TFRC
BAG6
UBL4A
Targeting to ER Membrane
Casein Kinase II
CSNK2A2
CSNK2A1
Nsp15
N
IGHV3-72
PPP1R7
SAE1
KEAP1
CNPY4
XPNPEP3
Orf3
RAD23B
RNH1
NME1
TPP2
MAPK3
MAPK1
ERK1/2
S
LGALS1
Glycosylation
B4GAT1
CHPF
ATL3
TNPO3
DNAJB9
ATP6AP2
SPTLC2
SCFD1
Autophagy
E
WLS
BRI3BP
IGHG4
DCP1A
TMEM43
ER Protein Degradation
PSMC4
NGLY1
USP28
USP13
ERLEC1
SEL1L
OS9
IGF2R
LATS1
VWA8
DNPEP
CEP72
NAGLU
Nsp16
SMAD2
WDR59
TRIM59
ACOT9
HSD17B4
Lipid Metabolism
Nuclear Import
KPNA2
KPNA6
KPNA1
KPNA5
KPNA3
KPNA4
NUP153
NUP50
C1QBP
Orf4b
MAN2B2
USP7
RNASEH1
Ubiquitination
TUBGCP2
DIPK2A
TTC27
YIPF6
LTN1
GCN1
M
TELO2
FANCI
Vesicle Transport
VAMP8
LMAN2L
LMBRD2
FUT8
CDS2
MARCH5
IPO11
NUP210
HEATR1
Nuclear Transport
MERS Viral Protein
Human Protein
Human-Human Protein Interaction
Virus-Human Protein Interaction (AP-MS)
Protein Complex
Biological Process

Figure S6. MERS-CoV Virus-Human Protein Interaction Network.

Virus-human protein- protein interaction map depicting high-confidence interactions (MiST ≥ 0.7 & Saint BFDR ≤0.05 & Average Spectral Counts ≥ 2) for MERS-CoV as derived from affinity purification-mass spectrometry (AP-MS). Viral bait proteins are depicted with yellow diamonds and human proteins with dark grey circles. Human-human interactions are depicted in thin, dark grey lines. Proteins within the same protein complexes or biological process are indicated with light yellow or light blue highlighting, respectively, and annotated accordingly. Human-human physical interactions, protein complex definitions, and biological process groupings are derived from CORUM (46), Gene Ontology (biological process), and manually curated from literature sources. Thin dashed grey lines are used to indicate the placement of node labels when adjacent node labels would have otherwise been obscured.

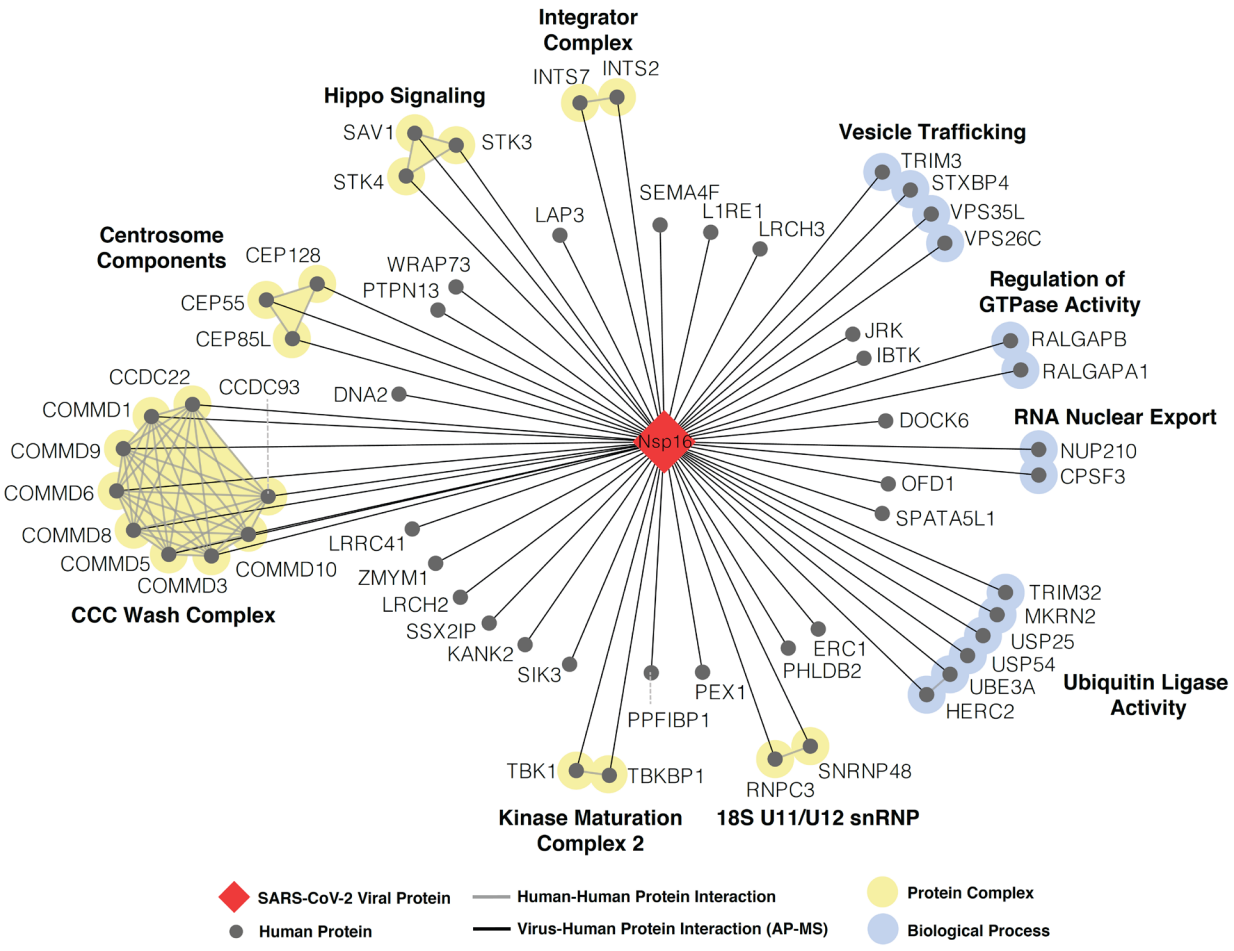


Figure S7. SARS-CoV-2 Nsp16 Virus-Host Protein Interaction Network.

Virus-human protein-protein interaction map depicting high-confidence interactions (MiST ≥ 0.7 & Saint BFDR ≤ 0.05 & Average Spectral Counts ≥ 2) for SARS-CoV-2 Nsp16 protein. This network is derived from affinity purification-mass spectrometry (AP-MS) data. Viral bait proteins are depicted with red diamonds and human proteins with dark grey circles. Human-human interactions are depicted in thin, dark grey lines. Proteins within the same protein complexes or biological process are indicated with light yellow or light blue highlighting, respectively, and annotated accordingly. Human-human physical interactions, protein complex definitions, and biological process groupings are derived from CORUM (46), Gene Ontology (biological process), and manually curated from literature sources.

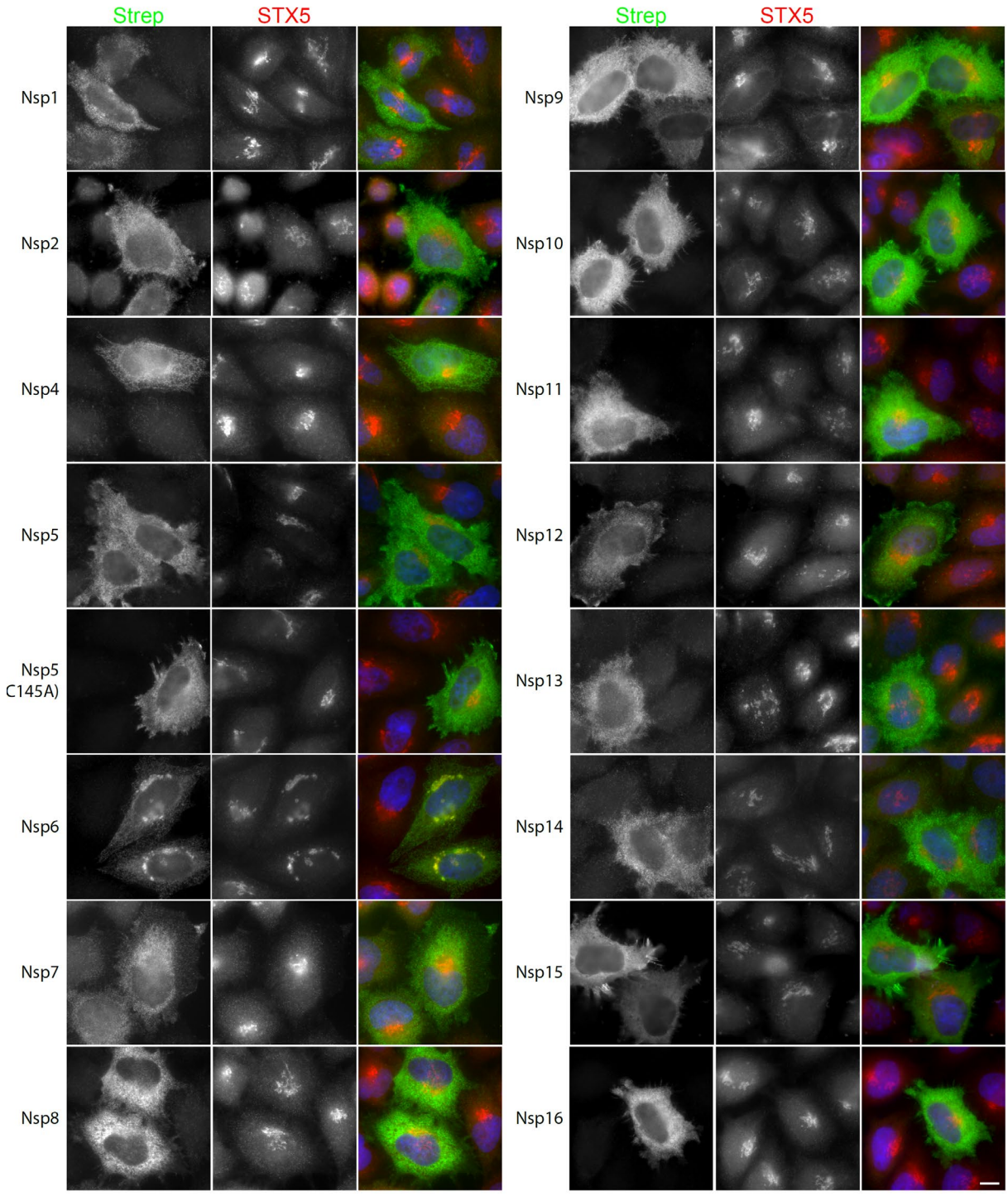


Figure S8. Immunolocalization of Strep-tagged SARS-CoV-2 non-structural (Nsp) proteins.

HeLaM cells were transfected with 2xStrep-tagged viral proteins, fixed, and immunostained with anti-Strep antibodies. Samples were also immunostained for the Golgi-localized protein Syntaxin 5 (STX5). Scale bar = 10 μm.

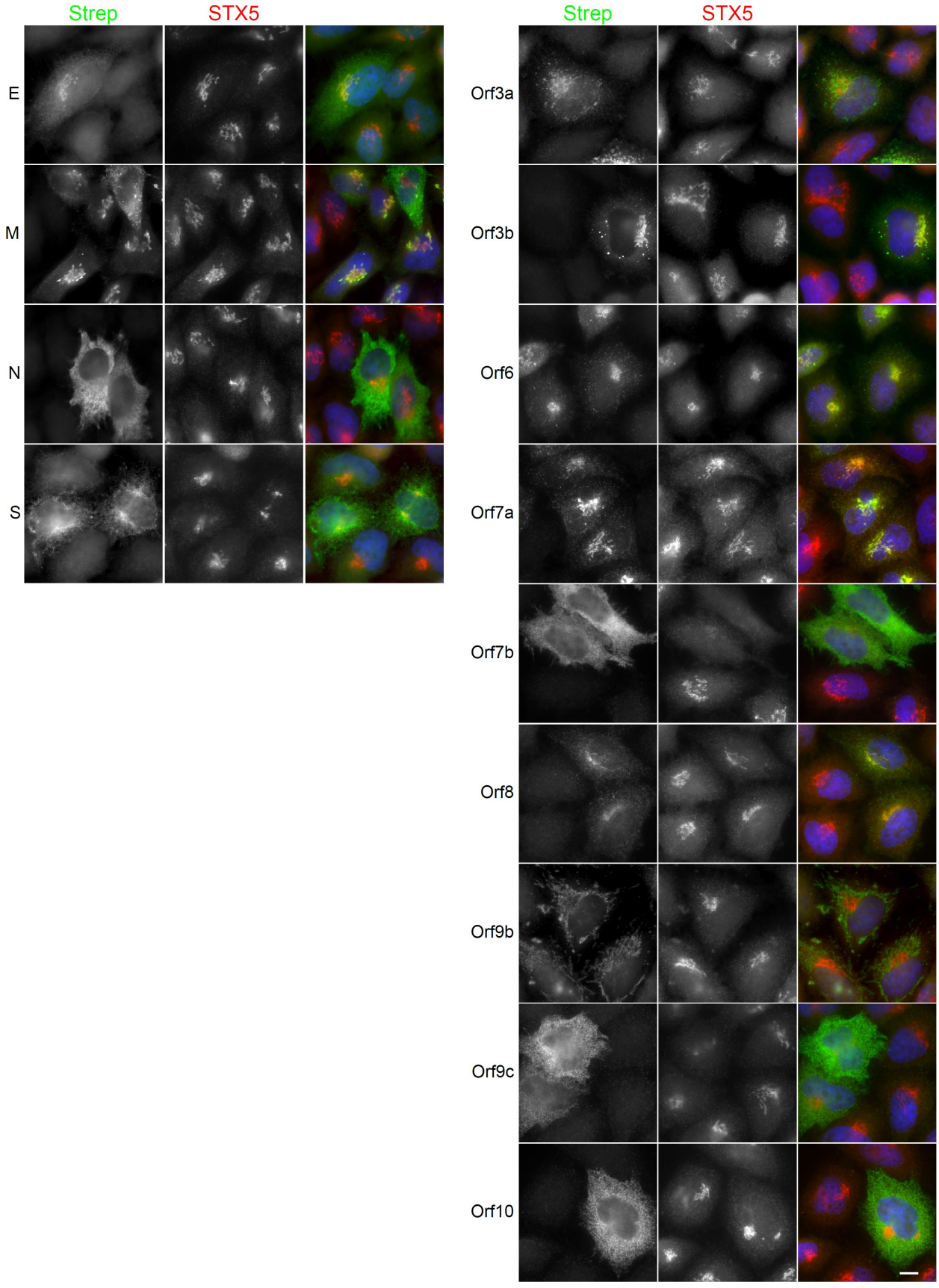

Strep
STX5
E
M
N
S
Strep
STX5
Orf3a
Orf3b
Orf6
Orf7a
Orf7b
Orf8
Orf9b
Orf9c
Orf10

Figure S9. Immunolocalization of Strep-tagged SARS-CoV-2 structural and accessory proteins.

HeLaM cells were transfected with 2xStrep-tagged viral proteins, fixed, and immunostained with anti-Strep antibodies. Samples were also immunostained for the Golgi- localized protein Syntaxin 5 (STX5). Scale bar = 10 μm.

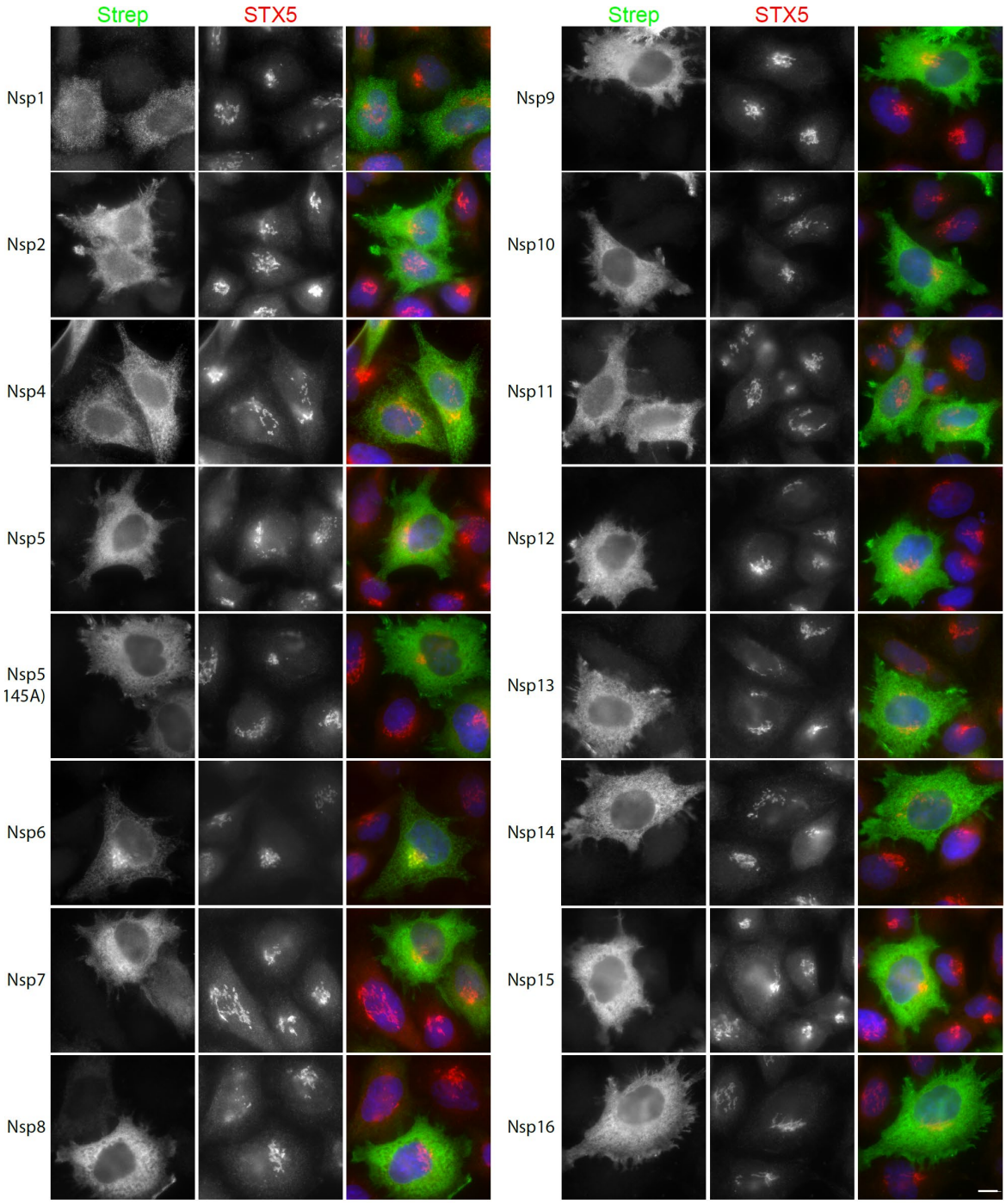


Figure S10. Immunolocalization of Strep-tagged SARS-CoV-1 non-structural (Nsp) proteins.

HeLaM cells were transfected with 2xStrep-tagged viral proteins, fixed, and immunostained with anti-Strep antibodies. Samples were also immunostained for the Golgi-localized protein Syntaxin 5 (STX5). Scale bar = 10 μm.

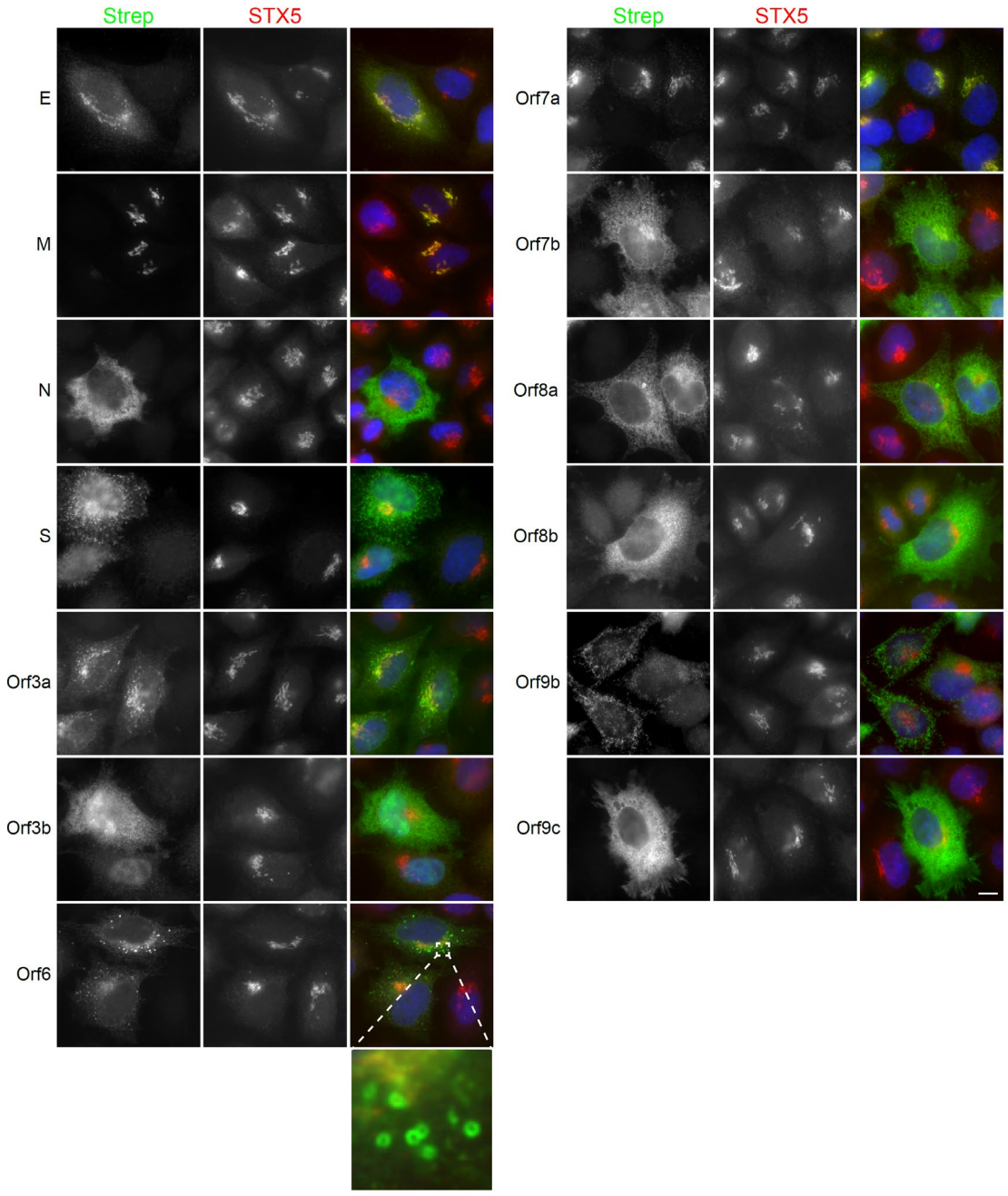


Figure S11. Immunolocalization of Strep-tagged SARS-CoV-1 structural and accessory proteins.

HeLaM cells were transfected with 2xStrep-tagged viral proteins, fixed, and immunostained with anti-Strep antibodies. Samples were also immunostained for the Golgi- localized protein Syntaxin 5 (STX5). Scale bar = 10 μm. Ring structures formed by SARS-CoV- 1 Orf6 highlighted in enlarged micrograph image.

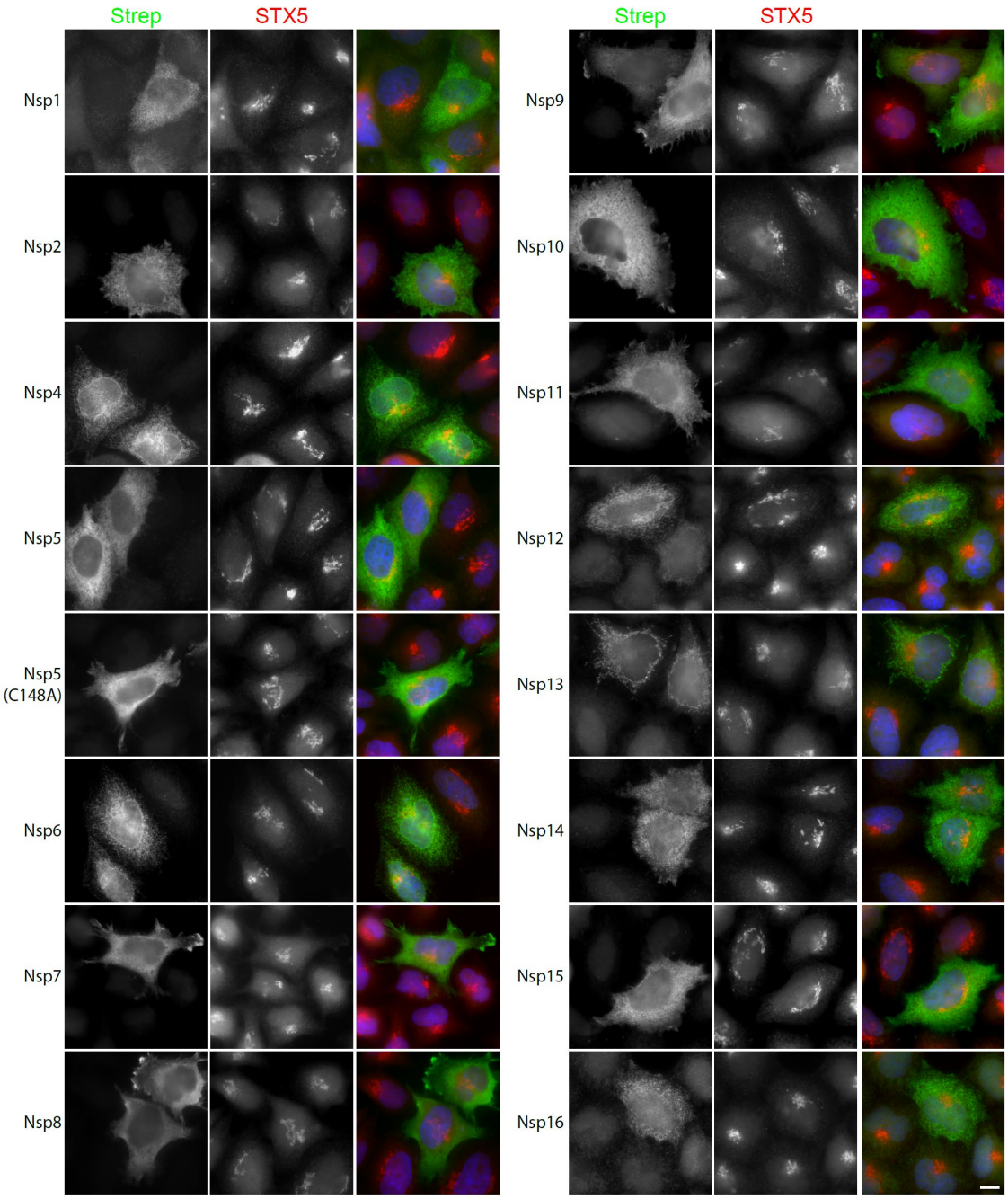


Figure S12. Immunolocalization of Strep-tagged MERS-CoV non-structural (Nsp) proteins. HeLaM cells were transfected with 2xStrep-tagged viral proteins, fixed, and immunostained with anti-Strep antibodies. Samples were also immunostained for the Golgi-localized protein Syntaxin 5 (STX5). Scale bar = 10 μm.

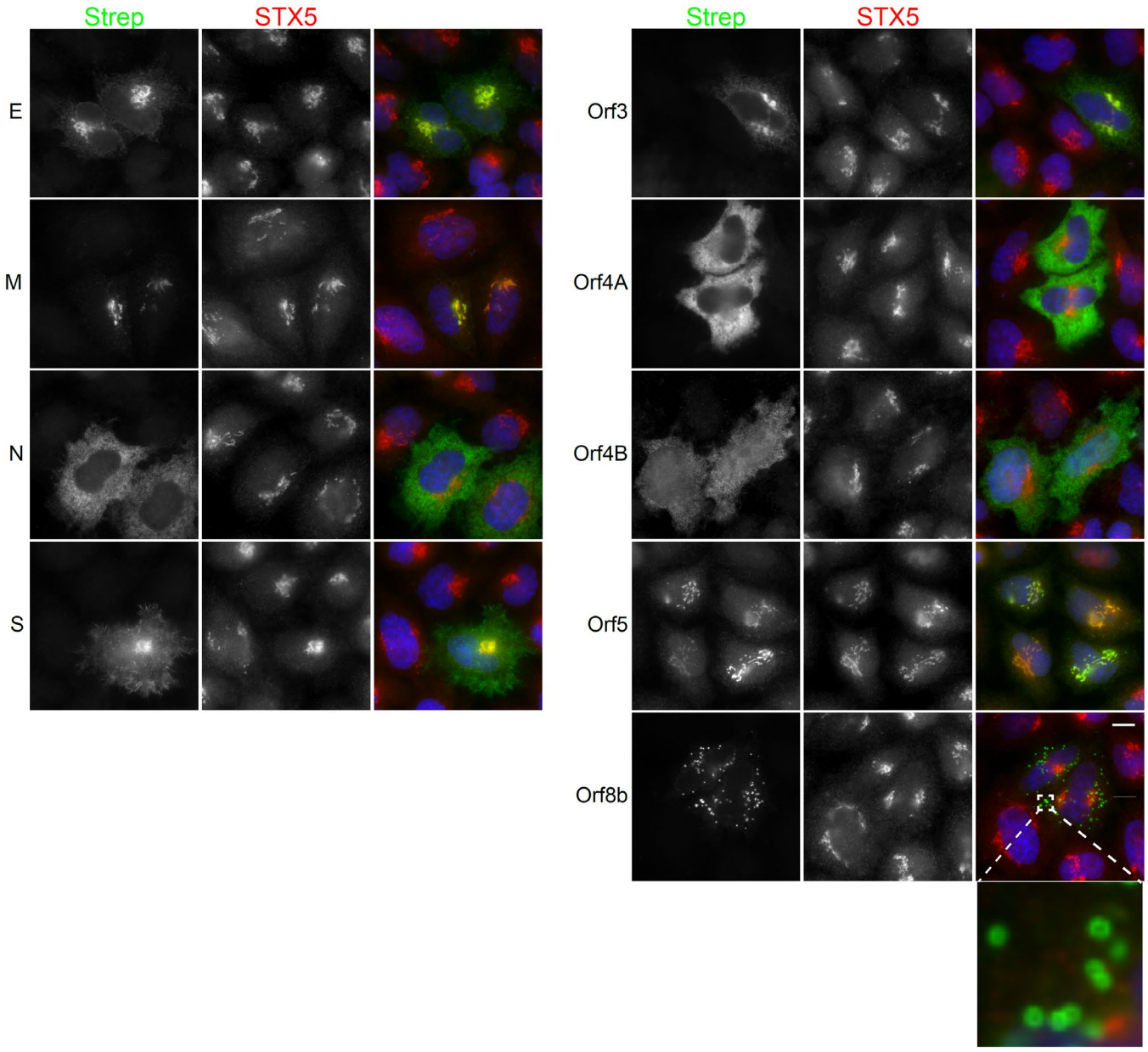


Figure S13. Immunolocalization of Strep-tagged MERS-CoV structural and accessory proteins.

HeLaM cells were transfected with 2xStrep-tagged viral proteins, fixed, and immunostained with anti-Strep antibodies. Samples were also immunostained for the Golgi- localized protein Syntaxin 5 (STX5). Scale bar = 10 μm. Ring structures formed by MERS-CoV Orf8b highlighted in enlarged micrograph image.

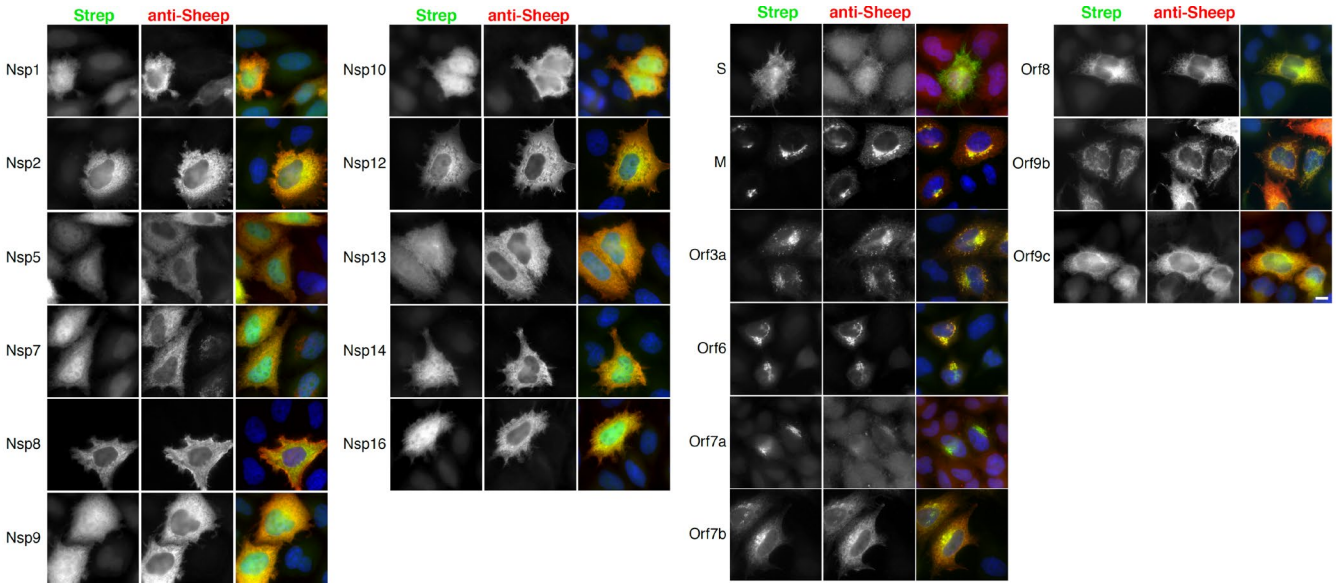


Figure S14. Validation of sheep anti-SARS-CoV-2 antibodies for immunohistochemistry.

HeLaM cells were transfected with the indicated 2xStrep-tagged viral proteins, fixed, and immunostained with anti-Strep antibodies and the appropriate anti-SARS-CoV-2 antiserum. Comparative visualization of anti-Strep and anti-sheep secondary antibodies are shown, with overlay at right. Sera against Nsp12, Nsp16, S, Orf7a, Orf8, and Orf9c did not perform well on infected cells and were excluded from further characterisation. Data generated with the 14 validated antisera are presented in Fig. 2B and quantified results as well as detailed analysis of the validation presented here is found in Table S3. Scale bar = 10 µm.

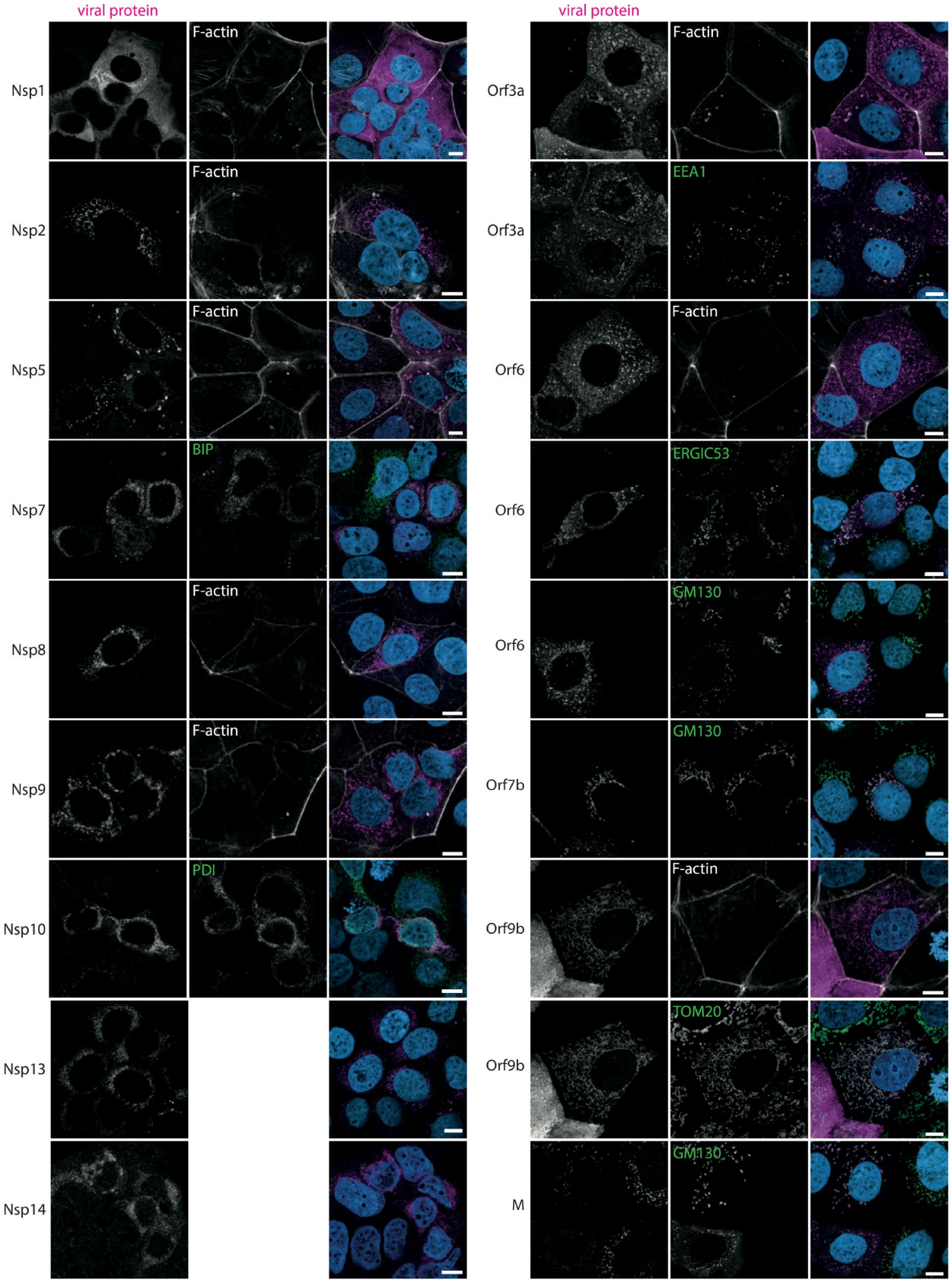

viral protein
F-actin
Nsp1
Nsp2
F-actin
Nsp5
F-actin
Nsp7
BIP
Nsp8
F-actin
Nsp9
F-actin
Nsp10
PDI
Nsp13
Nsp14
viral protein
F-actin
Orf3a
Orf3a
EEA1
Orf6
F-actin
Orf6
ERGIC53
Orf6
GM130
Orf7b
GM130
Orf9b
F-actin
Orf9b
TOM20
M
GM130

Figure S15. Immunolocalization of SARS-CoV-2 proteins in infected Caco-2 cells.

Caco-2 cells were infected with SARS-CoV-2 with a MOI of 0.1, fixed, and immunostained with specific polyclonal antibodies. To verify subcellular localization of viral proteins (magenta) after SARS- CoV-2 infection, counterstains were performed using Alexa Fluor 647-conjugated phalloidin or the following markers (green, Alexa Fluor 488): for endoplasmatic reticulum BiP (binding immunoglobulin protein) and PDI (protein disulfide isomerase), for endosomes EEA1 (early endosome antigen 1), for Golgi GM130 (Golgin subfamily A member 2), for ER-Golgi intermediate compartment ERGIC53, for mitochondria Tom20 (translocase of the outer membrane of mitochondria). Nuclei were stained with DAPI. Scale bar = 10 µm.

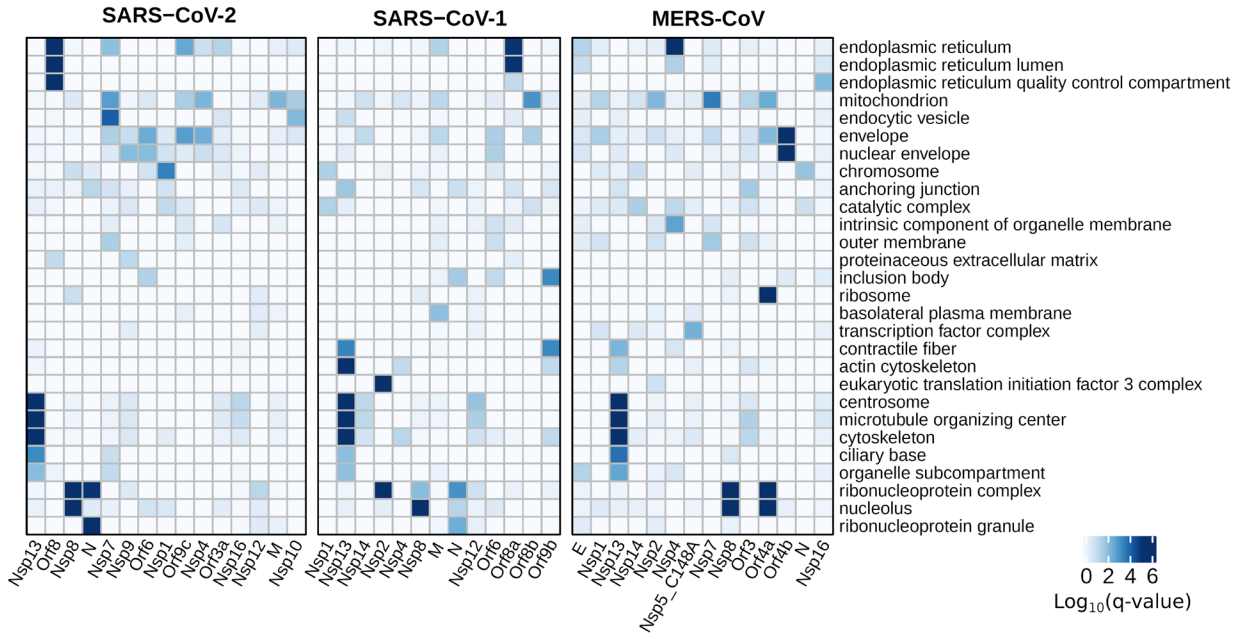


Figure S16. Enrichment of cellular compartment annotations for the interactors of each viral protein.

Gene Ontology (GO) enrichment analysis of cellular compartment annotations of the high-confidence interactors of the three viruses. Only some of the top most significant terms are included per virus. Color indicates -log10(q).

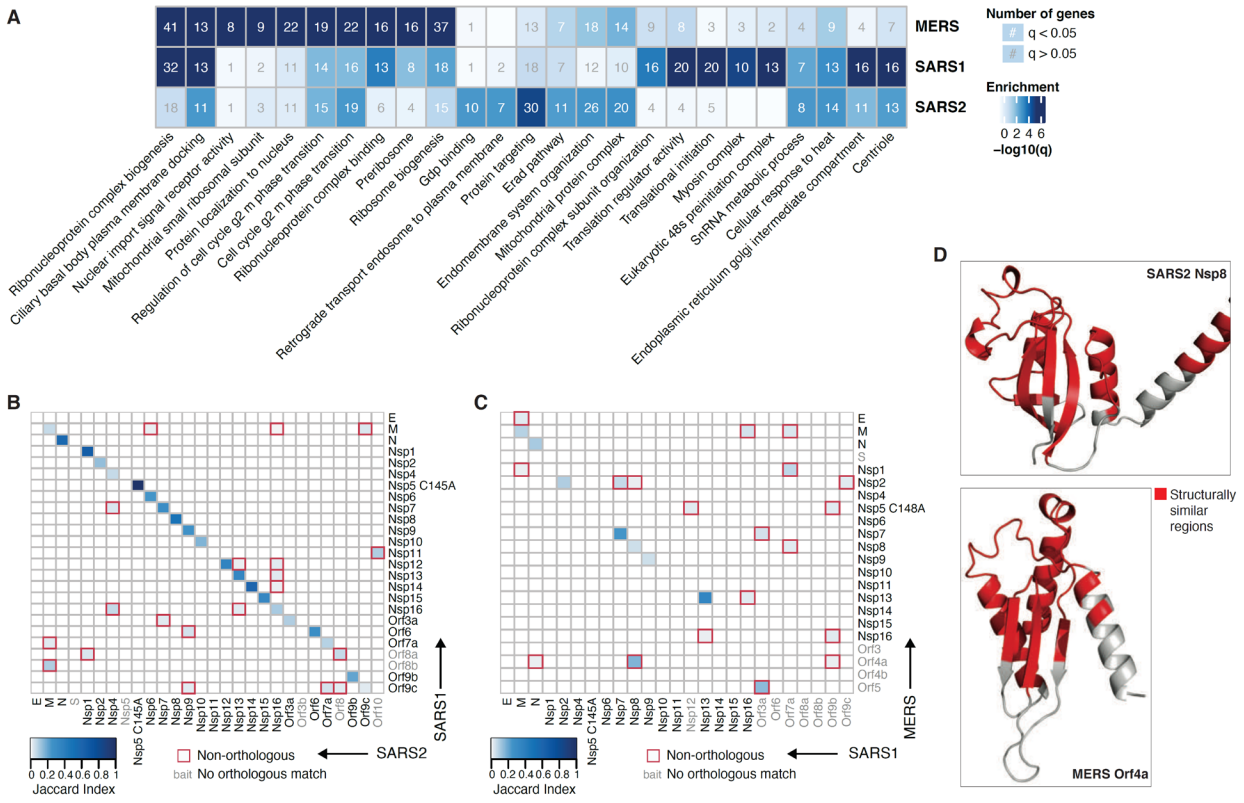


Figure S17. Comparison of enriched terms and shared interactors across viruses.

A) Gene Ontology (GO) enrichment analysis of the high-confidence interactors of the three viruses. The top ten most significant terms are included per virus. Color indicates -log10(q). Number indicates number of genes; white numbers denote significant enrichment (q<0.05), whereas grey numbers indicate non-significance (q>0.05). **B**) Heatmap depicting overlap in protein-protein interactions (Jaccard index) between all baits from SARS-CoV-1 and SARS-CoV-2. Baits in grey were not assessed, do not exist, or do not have high-confidence interactors in the alternate virus. Non- orthologous baits are highlighted with a red square. **C**) Heatmap depicting overlap in protein- protein interactions (Jaccard index) between all baits from SARS-CoV-1 and MERS-CoV. Baits in grey were not assessed, do not exist, or do not have high-confidence interactors in the alternate virus. Non-orthologous baits are highlighted with a red square. **D**) The structure of the C-terminal region of SARS-CoV-2 Nsp8 (upper panel) and a predicted structural model of MERS-CoV Orf4a (lower panel). Red represents structurally similar regions as determined by Geometricus.

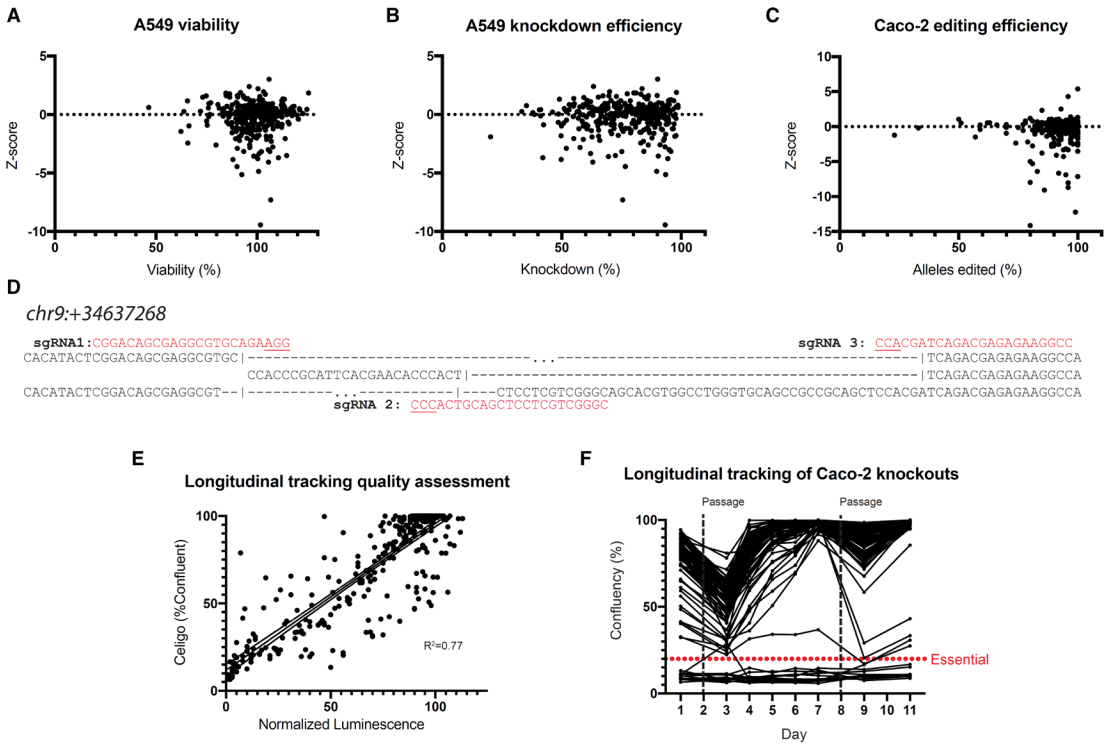


Figure S18. Viabilities, knockdown efficiencies and editing efficiencies in response to siRNA and CRISPR perturbations.

A) Z-score plotted against viability in A549-ACE2 siRNA knockdowns. **B**) Z-score plotted against siRNA knockdown efficiency in A549-ACE2 cells for 327 of the 332 genes included in the final siRNA dataset. Knockdown efficiency was not obtained for the remaining 5 genes. **C**) Z-score plotted against editing efficiency (ICE-D score) for 227 of the 288 genes included in the final Caco-2 CRISPR dataset. ICE-D scores were not obtained for the remaining 61 genes. **D**) Representative genotype in Caco-2 SIGMAR1 knockout. Use of multiguide strategy causes genomic dropout between sgRNAs. Plurality of alleles at SIGMAR1 locus have undergone frameshift mutation. **E**) Correlation between quantitative but destructive measurement of cell viability using CellTiter-Glo and non-invasive longitudinal tracking using brightfield imaging. Both measurements are in agreement suggesting both methods can be used to determine gene essentiality (error bars ±1 S.D., R2=0.77). These data are from a separate experiment using A549 cells. **F**) Longitudinal tracking of Caco-2 gene knockout pools using brightfield imaging. Pools were imaged every day for 11 days except for days of passaging (days 2 and 8, vertical dotted line). The majority of pools showed exponential growth. However, several stayed below the limit of detection (red horizontal line) suggesting pools were lost due to the essential nature of the gene.

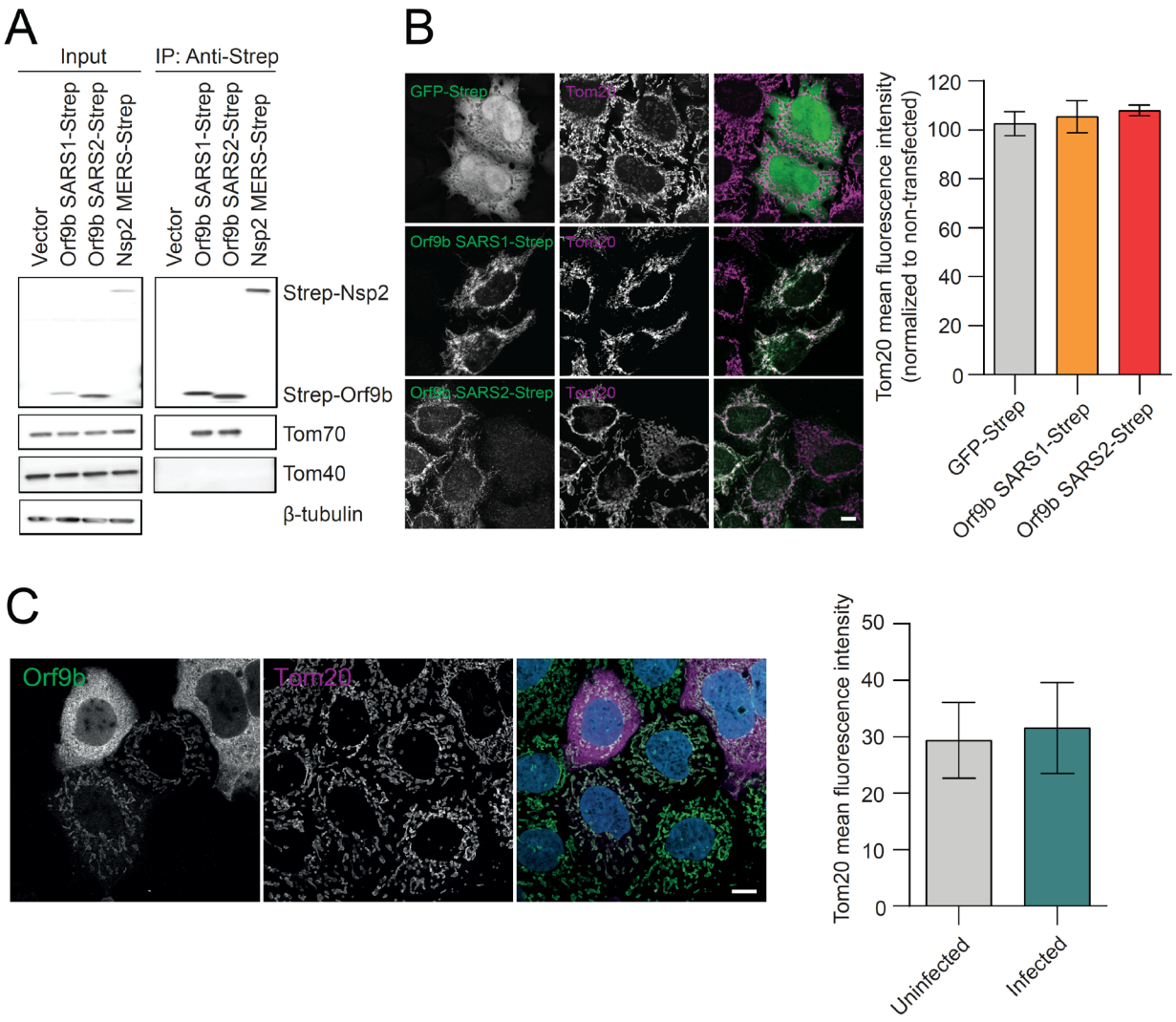


Figure S19. Orf9b interacts specifically with Tom70.

A) Co-immunoprecipitation between Strep-Orf9b and endogenous Tom70. A549 cells were transfected with Strep-tagged Orf9b from SARS-CoV-1 and SARS-CoV-2 along with Nsp2 from MERS-CoV. IP was performed using anti- Strep beads and representative immunoblots of whole cell lysates and eluates are shown. **B**) Immunostained images of SARS-CoV-2 Orf9b-expressing HeLaM cells stained for Tom20 and Strep-Orf9b (left). Mean fluorescence intensity ± SD values of Tom20 in GFP-Strep and Orf9b expressing cells (normalized to non-transfected cells; right). **C**) Representative immunostained images of Orf9b and Tom20 upon SARS-CoV-2 infection. IP = immunoprecipitation; SD = standard deviation.

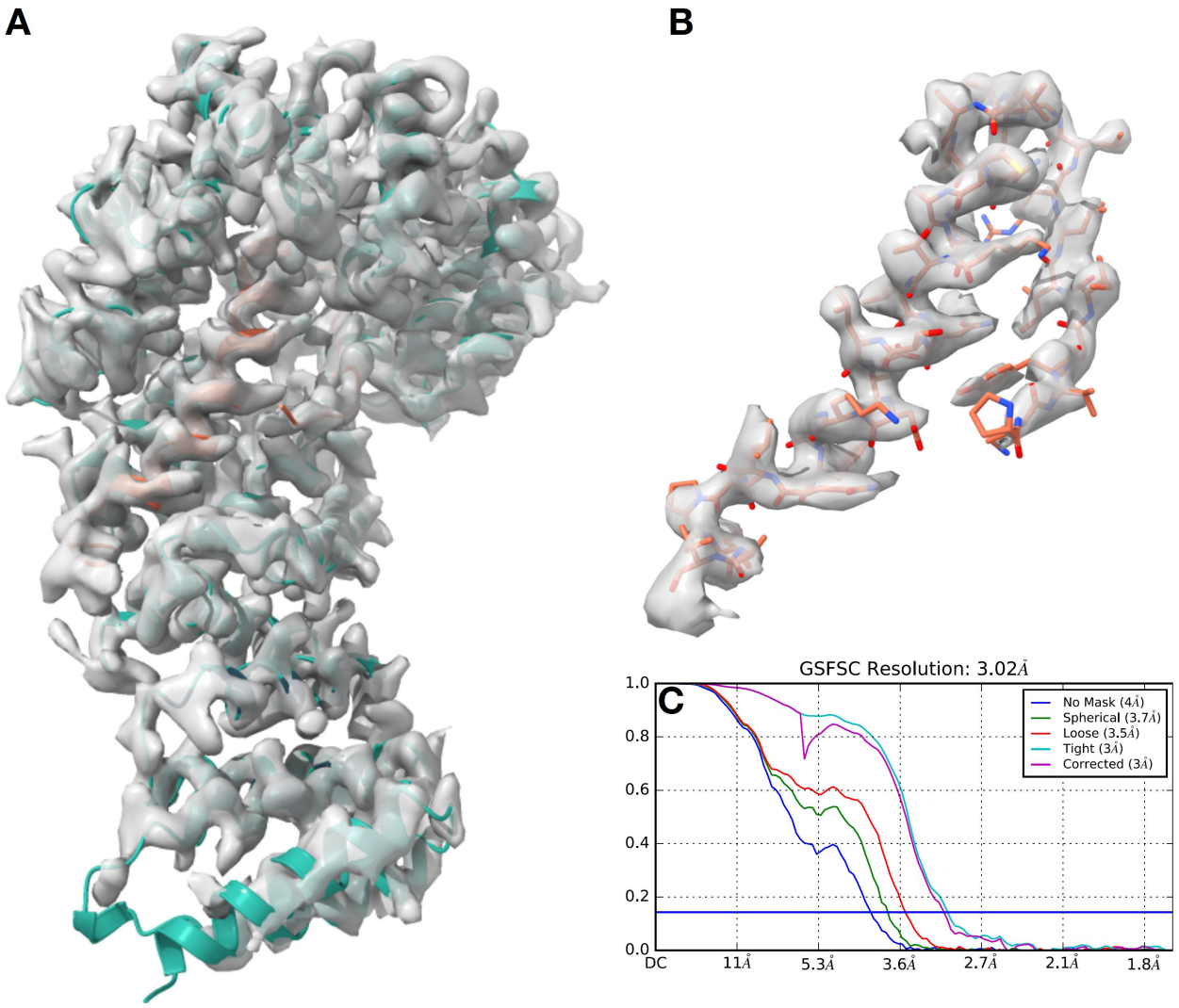


Figure S20. Orf9b-Tom70 cryoEM density map and the Fourier Shell Correlation of the final reconstruction.

A) CryoEM density (weighted by FSC and sharpened with a B-factor of -145) of Orf9b-Tom70 complex with the built atomic models depicted as ribbon. Tom70 is in green, Orf9b is in orange. **B**) Magnified view of the cryoEM density just around Orf9b indicated in sticks showing a good agreement between the density and the model. **C**) Gold standard Fourier shell correlation of the resulting reconstruction as output by cryosparc software package.

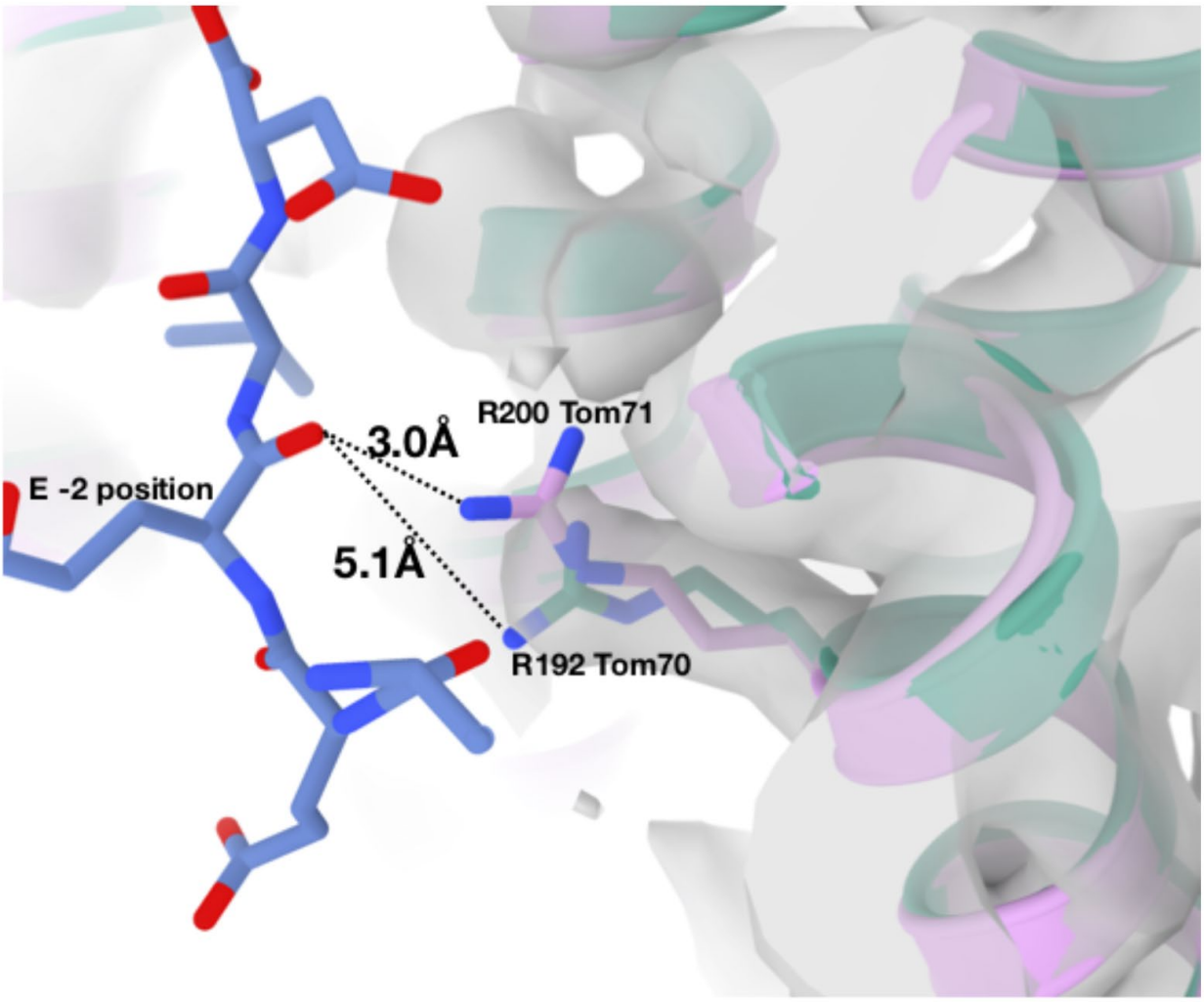


Figure S21. Subtle conformational changes at the MEEVD binding site of Tom70.

Magnified view of R192/R200 (human Tom70/yeast Tom71), which is a key interacting residue with the EEVD motif from Hsp70/Hsp90. The conformation in yeast Tom71 [competent to bind EEVD, PDB:3FP2 (101)] is shown in lavender. Conformation in our human Tom70 structure is shown in green, indicating that the arginine (R) is moved out of position to hydrogen bond with the glutamate. The EEVD peptide is shown as sticks in blue with the E at the -2 position (where terminal D is position 0) indicated. The cryoEM density is also shown depicting good agreement between the model and the density for R192.

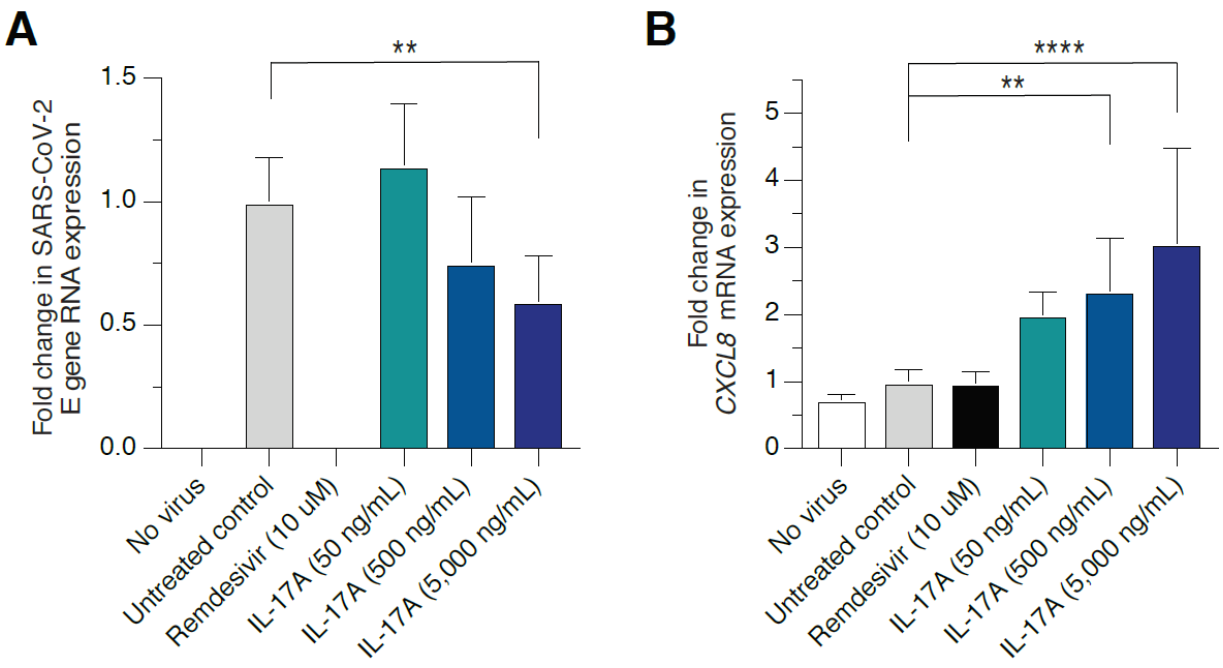


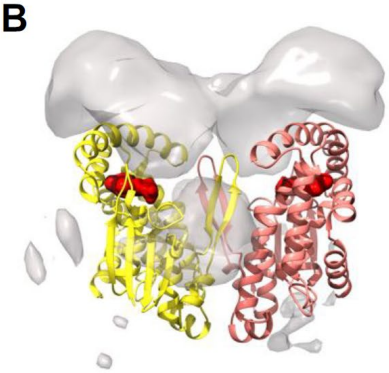
Figure S22. IL-17A treatment of A549-ACE2 cells concurrent with SARS-CoV-2 inoculation causes decreases in viral replication.

A) Viral gene E RNA expression 24 hours after infection and treatment with indicated agents in A549-ACE2 cells. **B**) CXCL8 mRNA expression 24 hours after infection and treatment with indicated agents in A549-ACE2 cells. Plots represent 2 biological replicates with 3 technical replicates each. * = p <0.05, ** = p < 0.005, **** = p < 0.00005, one-way ANOVA relative to untreated control condition with Dunnet multiple comparison correction. Error bars indicate standard deviation.

A

Method	# Models
ClusPro	98
ZDock	10
HDock	100
Gramm-X	100
SwarmDock	19
PatchDock	93

B



C

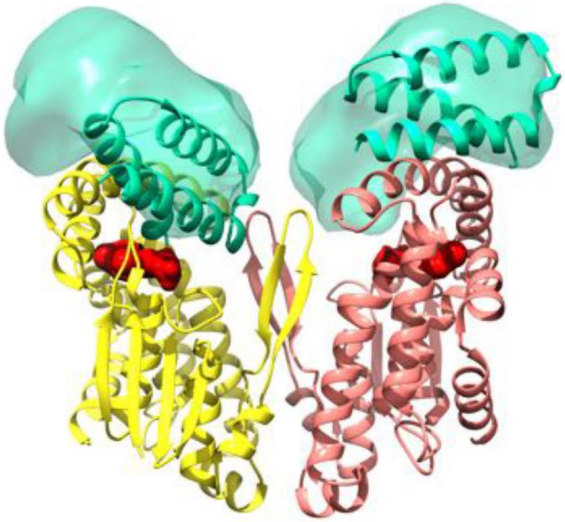


Figure S23. Predicted docking modes of PGES-2 and Nsp7.

A) The structure of Nsp7 was docked against a homology model of the PGES-2 dimer (yellow and pink) using a number of docking programs. The number of good scoring models produced by each docking protocol is shown. **B**) The combined localization density of all 420 good scoring models. **C**) The top two clusters of solutions (cyan volume) are symmetry-related and localize to the lobe of mPGES-2 adjacent to the indomethacin binding site (red). Ribbon models of the top scoring models from PatchDock (left) and ZDock (right) represent the two distinct binding modes contained in this cluster of solutions.

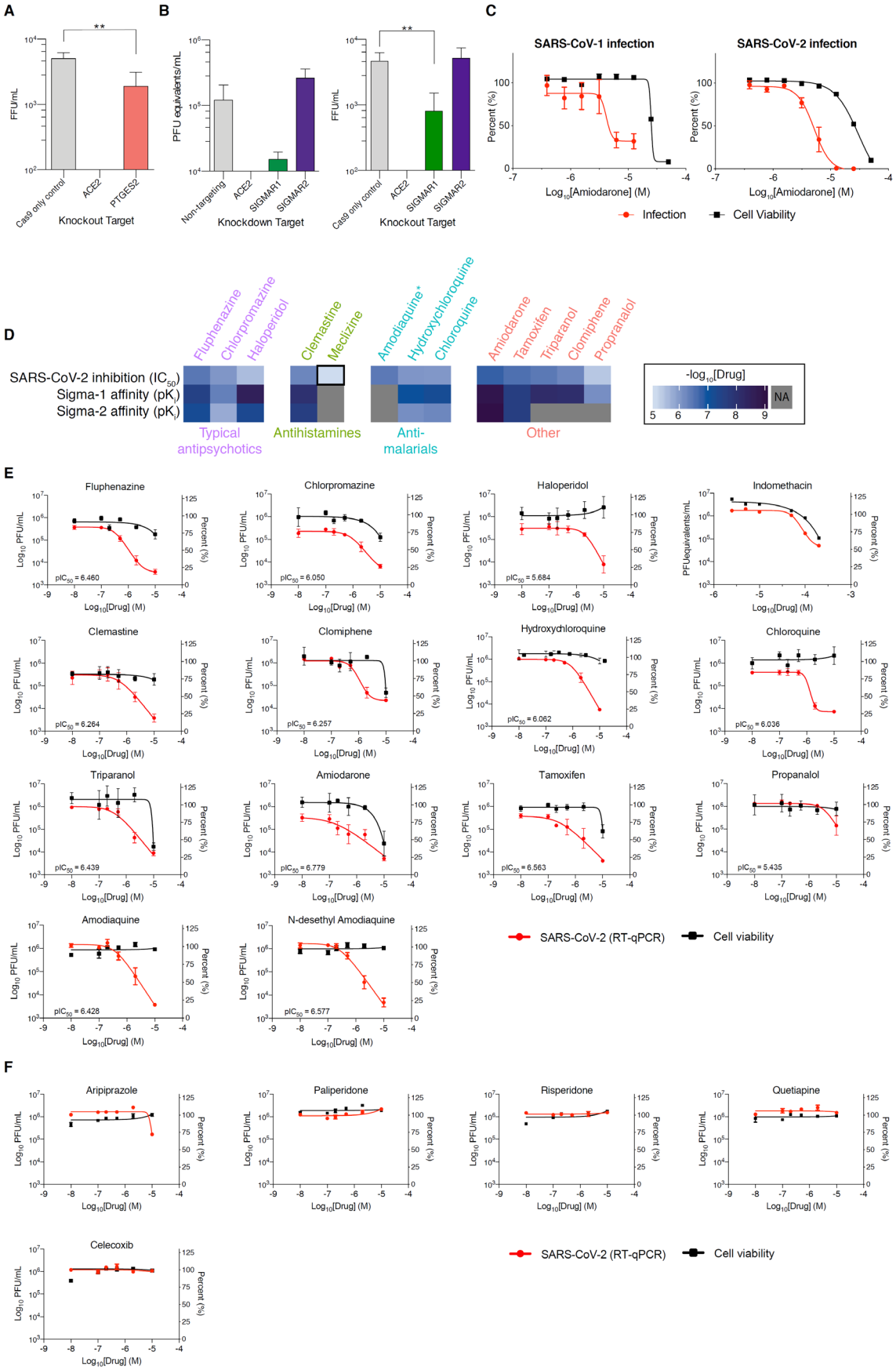

A
B
C
FFU/mL
Cas9 only control
ACE2
PTGES2
Knockout Target
PFU equivalents/mL
Non-targeting
SIGMAR1
SIGMAR2
Knockdown Target
SARS-CoV-1 infection
SARS-CoV-2 infection
Percent (%)
Log10[Amiodarone] (M)
Infection
Cell Viability
D
Fluphenazine
Chlorpromazine
Haloperidol
Clemastine
Meclizine
Amodiaquine*
Hydroxychloroquine
Chloroquine
Amiodarone
Tamoxifen
Triparanol
Clomiphene
Propranalol
SARS-CoV-2 inhibition (IC50)
Sigma-1 affinity (pKi)
Sigma-2 affinity (pKi)
-log10[Drug]
NA
Typical antipsychotics
Antihistamines
Anti-malarials
Other
E
Fluphenazine
Chlorpromazine
Haloperidol
Indomethacin
Clemastine
Clomiphene
Hydroxychloroquine
Chloroquine
Triparanol
Amiodarone
Tamoxifen
Propanalol
Amodiaquine
N-desethyl Amodiaquine
SARS-CoV-2 (RT-qPCR)
Cell viability
F
Aripiprazole
Paliperidone
Risperidone
Quetiapine
Celecoxib
SARS-CoV-2 (RT-qPCR)
Cell viability

Figure S24. Perturbation of drug targets and performance of selected drugs against coronavirus replication in vitro.

A) SARS-CoV-2 replication in Caco-2 cells after knockout of PTGES2 or controls. ** = $p < 0.005$, unpaired t test. **B**) SARS-CoV-2 replication in A549-ACE2 cells or Caco-2 cells after knockdown and knockout, respectively, of SIGMAR1, SIGMAR2 (TMEM97) or controls. ** = $p < 0.005$, one-way ANOVA relative to control with Dunnet multiple comparison correction. **C**) Antiviral activity of amiodarone against SARS-CoV-1 (left) and SARS- CoV-2 (right) in Vero E6 cells. **D**) Clinically-approved sigma receptor-targeting drugs with verified anti-SARS-CoV-2 activity by clinical drug class. Heatmap indicates, from top to bottom: pIC50 = (-log10[IC50]) of the drug against SARS-CoV-2; reported pKi = [-log10(Ki)] of the drug against the sigma-1 receptor of any species; reported pKi of the drug against sigma- 2 receptor of any species. SARS-CoV-2 IC50 was determined in A549-ACE2 cells or in Vero E6 cells where indicated by a black border. Grey boxes indicate no value was found reported in the literature. Amodiaquine has been taken off the American market and is indicated by an asterisk. **E**) Performance of putative drugs for repurposing discussed in the paper against SARS-CoV-2 in vitro in A549-ACE2 cells. Error bars indicate standard deviation. **F**) Performance of clinical comparator drugs discussed in the paper against SARS-CoV-2 in vitro in A549-ACE2 cells. Error bars indicate standard deviation. Quantification of the data in D-F is found in Table S10.

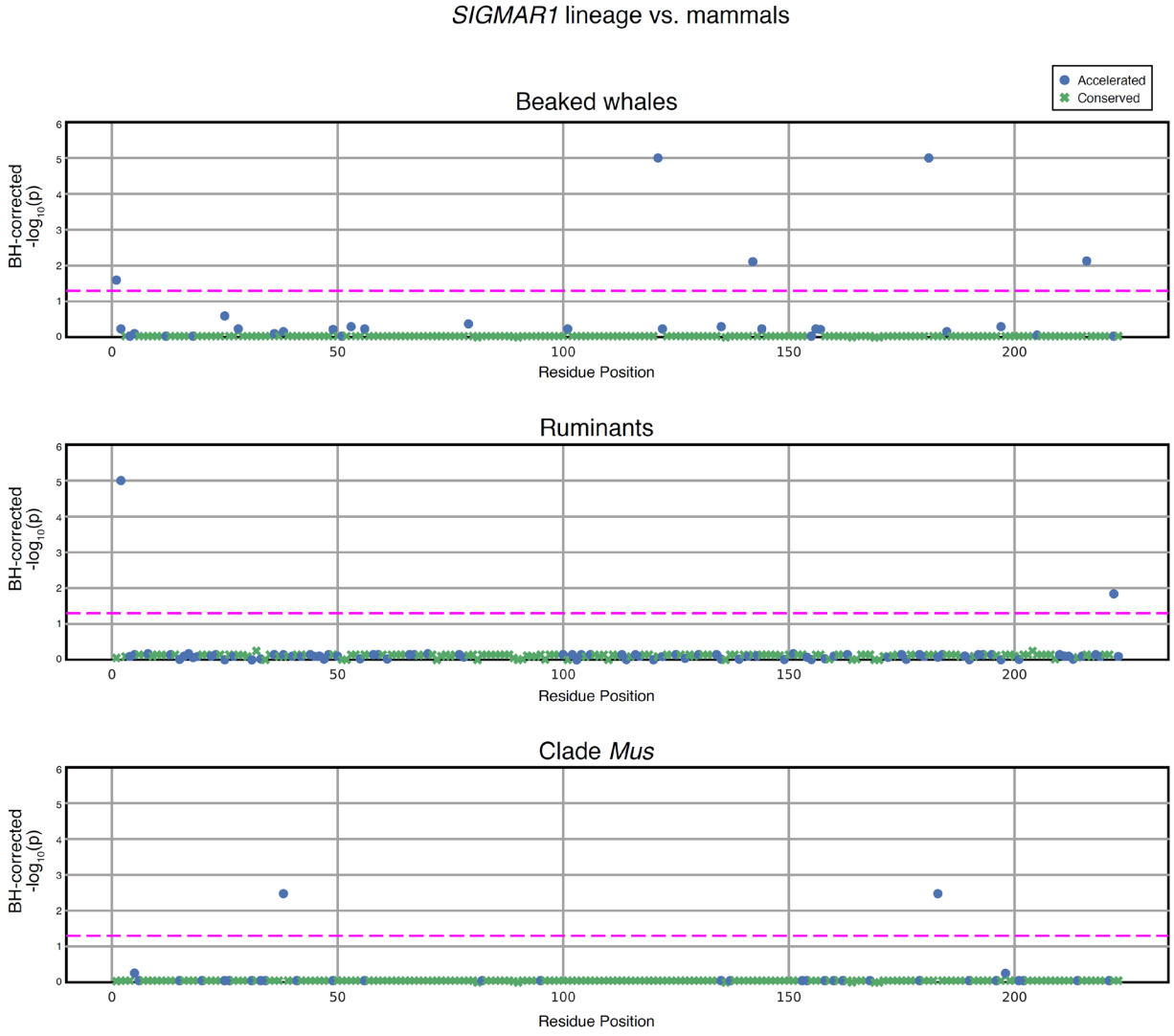


Figure S25. Departures from neutral evolution in SIGMAR1.

Benjamini-Hochberg-corrected p-values (y-axis) for accelerated (blue circles) or conserved (green Xs) evolution at codons in SIGMAR1 in the denoted lineages relative to the neutral rate in mammals.

Table S1. Protein sequence similarity. Protein sequence similarities for SARS-CoV-2:SARS-CoV-1, SARS-CoV-2:MERS, SARS-CoV-1:SARS-CoV-2 and MERS:SARS-CoV-2. Identitiesare only presented for genes with the same name in the compared virus pairs.

Table S2. Unthresholded and high-confidence virus-host protein interactions. Contains unthresholded (SARS-CoV-2_Unthresholded, SARS-CoV-1_Unthresholded, and MERS-CoV_Unthresholded tabs) as well as high-confidence (SARS-CoV-2_HighConfidence, SARS-CoV-1_HighConfidence, and MERS-CoV_HighConfidence tabs) virus-host protein interaction data from affinity purification-mass spectrometry (AP-MS) studies. Column descriptions are indicated in the final tab.

Table S3. Information on localization of viral proteins derived from immunolocalization experiments and predictions. The localization information for each of the viral proteins studied by immunostaining and sequence based predictions. It also includes information on use and validation of the anti-SARS-CoV-2 protein anti-sera.

Table S4. GO enrichment analysis results. Includes significantly enriched GO terms associated with each cluster and high-confidence interactors of each virus.

Table S5. Clustering and differential interaction scores. Contains virus-host interactions defined as high confidence for at least one virus. Only interactions involving viral bait proteins shared across all three viruses are considered. Differential interaction scores for all comparisons are included (SARS-CoV-1:MERS-CoV, SARS-CoV-2:MERS-CoV, SARS-CoV-2:SARS-CoV-1, SARS-CoV-1-and-2:MERS-CoV). MiST scores, Saint scores (Saint score, BFDR, FoldChange), and average spectral counts for each interaction are also included.

Table S6. A549-ACE2 siRNA knockdown data. Includes averaged Z-scores, unaveraged Z-scores, viabilities and knockdown efficiencies.

Table S7. Caco-2 CRISPR knockout data. Includes averaged Z-scores, unaveraged Z-scores and editing efficiencies.

Table S8. Statistics of the Orf9b-Tom70 cryoEM structure presented in this work.

Table S9. Statistics relating to genetic analysis of soluble IL17RA in COVID Host Genetics Initiative data.

Table S10. Sourcing of drugs for in vitro analysis with literature affinity values and measured SARS-CoV-2 IC50 values.

Table S11. Drug list, matching criteria, outcomes and comparative data for real-world comparison of host factor-targeting drugs.

Table S12. siRNA library sequences.

Table S13. siRNA library qPCR primers.

Table S14. sgRNA library sequences.

Table S15. Cas9 knockout amplicon PCR and sequencing primers

Movie S1. Conformational change between yeast Tom71 (without substrate) and Orf9b-bound human Tom70 focusing on the core domain. Movie starts with just the core domain of yeast Tom71 substrate free structure [PDB:3FP3(101)] shown as white ribbon and then interpolates the core domain to our Orf9b-Tom70 structure (Tom70 in green, Orf9b in orange).

References and Notes

1. J. Liu, W. Xie, Y. Wang, Y. Xiong, S. Chen, J. Han, Q. Wu, A comparative overview of COVID-19, MERS and SARS: Review article. Int. J. Surg. 81, 1–8 (2020). doi:10.1016/j.ijsu.2020.07.032

2. J. H. Beigel, K. M. Tomashek, L. E. Dodd, A. K. Mehta, B. S. Zingman, A. C. Kalil, E. Hohmann, H. Y. Chu, A. Luetkemeyer, S. Kline, D. Lopez de Castilla, R. W. Finberg, K. Dierberg, V. Tapson, L. Hsieh, T. F. Patterson, R. Paredes, D. A. Sweeney, W. R. Short, G. Touloumi, D. C. Lye, N. Ohmagari, M.-d. Oh, G. M. Ruiz-Palacios, T. Benfield, G. Fätkenheuer, M. G. Kortepeter, R. L. Atmar, C. B. Creech, J. Lundgren, A. G. Babiker, S. Pett, J. D. Neaton, T. H. Burgess, T. Bonnett, M. Green, M. Makowski, A. Osinusi, S. Nayak, H. C. Lane, ACTT-1 Study Group Members, Remdesivir for the treatment of Covid-19—Final report. N. Engl. J. Med. 383, 1813–1826 (2020). doi:10.1056/NEJMoa2007764

3. RECOVERY Collaborative Group, Dexamethasone in Hospitalized Patients with Covid-19-Preliminary Report. N. Engl. J. Med. 10.1056/nejmoa2021436 (2020). doi:10.1056/nejmoa2021436

4. M. Becerra-Flores, T. Cardozo, SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. Int. J. Clin. Pract. 74, e13525 (2020). doi:10.1111/ijcp.13525

5. D. E. Gordon, G. M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K. M. White, M. J. O'Meara, V. V. Rezelj, J. Z. Guo, D. L. Swaney, T. A. Tummino, R. Hüttenhain, R. M. Kaake, A. L. Richards, B. Tutuncuoglu, H. Foussard, J. Batra, K. Haas, M. Modak, M. Kim, P. Haas, B. J. Polacco, H. Braberg, J. M. Fabius, M. Eckhardt, M. Soucheray, M. J. Bennett, M. Cakir, M. J. McGregor, Q. Li, B. Meyer, F. Roesch, T. Vallet, A. Mac Kain, L. Miorin, E. Moreno, Z. Z. C. Naing, Y. Zhou, S. Peng, Y. Shi, Z. Zhang, W. Shen, I. T. Kirby, J. E. Melnyk, J. S. Chorba, K. Lou, S. A. Dai, I. Barrio-Hernandez, D. Memon, C. Hernandez-Armenta, J. Lyu, C. J. P. Mathy, T. Perica, K. B. Pilla, S. J. Ganesan, D. J. Saltzberg, R. Rakesh, X. Liu, S. B. Rosenthal, L. Calviello, S. Venkataramanan, J. Liboy- Lugo, Y. Lin, X.-P. Huang, Y. Liu, S. A. Wankowicz, M. Bohn, M. Safari, F. S. Ugur, C. Koh, N. S. Savar, Q. D. Tran, D. Shengjuler, S. J. Fletcher, M. C. O'Neal, Y. Cai, J. C. J. Chang, D. J. Broadhurst, S. Klippsten, P. P. Sharp, N. A. Wenzell, D. Kuzuoglu-Ozturk, H.-Y. Wang, R. Trenker, J. M. Young, D. A. Cavero, J. Hiatt, T. L. Roth, U. Rathore, A. Subramanian, J. Noack, M. Hubert, R. M. Stroud, A. D. Frankel, O. S. Rosenberg, K. A. Verba, D. A. Agard, M. Ott, M. Emerman, N. Jura, M. von Zastrow, E. Verdin, A. Ashworth, O. Schwartz, C. d'Enfert, S. Mukherjee, M. Jacobson, H. S. Malik, D. G. Fujimori, T. Ideker, C. S. Craik, S. N.

Floor, J. S. Fraser, J. D. Gross, A. Sali, B. L. Roth, D. Ruggero, J. Taunton, T. Kortemme, P. Beltrao, M. Vignuzzi, A. García-Sastre, K. M. Shokat, B. K. Shoichet, N. J. Krogan, A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 583, 459–468 (2020). doi:10.1038/s41586-020- 2286-9

6. G. Teo, G. Liu, J. Zhang, A. I. Nesvizhskii, A.-C. Gingras, H. Choi, SAINTexpress: Improvements and additional features in Significance Analysis of INTeractome software. J. Proteomics 100, 37–43 (2014). doi:10.1016/j.jprot.2013.10.023

7. S. Jäger, P. Cimermancic, N. Gulbahce, J. R. Johnson, K. E. McGovern, S. C. Clarke, M. Shales, G. Mercenne, L. Pache, K. Li, H. Hernandez, G. M. Jang, S. L. Roth, E. Akiva, J. Marlett, M. Stephens, I. D'Orso, J. Fernandes, M. Fahey, C. Mahon, A. J. O'Donoghue, A. Todorovic, J. H. Morris, D. A. Maltby, T. Alber, G. Cagney, F. D. Bushman, J. A. Young, S. K. Chanda, W. I. Sundquist, T. Kortemme, R. D. Hernandez, C. S. Craik, A. Burlingame, A. Sali, A. D. Frankel, N. J. Krogan, Global landscape of HIV-human protein complexes. Nature 481, 365–370 (2011). doi:10.1038/nature10719

8. J. C. Young, N. J. Hoogenraad, F. U. Hartl, Molecular chaperones Hsp90 and Hsp70 deliver preproteins to the mitochondrial import receptor Tom70. Cell 112, 41–50 (2003). doi:10.1016/S0092-8674(02)01250-3

9. R. Lin, S. Paz, J. Hiscott, Tom70 imports antiviral immunity to the mitochondria. Cell Res. 20, 971–973 (2010). doi:10.1038/cr.2010.113

10. B. Wei, Y. Cui, Y. Huang, H. Liu, L. Li, M. Li, K.-C. Ruan, Q. Zhou, C. Wang, Tom70 mediates Sendai virus-induced apoptosis on mitochondria. J. Virol. 89, 3804–3818 (2015). doi:10.1128/JVI.02959-14

11. A. M. Edmonson, D. K. Mayfield, V. Vervoort, B. R. DuPont, G. Argyropoulos, Characterization of a human import component of the mitochondrial outer membrane, TOMM70A. Cell Commun. Adhes. 9, 15–27 (2002). doi:10.1080/15419060212186

12. M. J. Baker, A. E. Frazier, J. M. Gulbis, M. T. Ryan, Mitochondrial protein-import machinery: Correlating structure with function. Trends Cell Biol. 17, 456–464 (2007). doi:10.1016/j.tcb.2007.07.010

13. J. Brix, K. Dietmeier, N. Pfanner, Differential recognition of preproteins by the purified cytosolic domains of the mitochondrial import receptors Tom20, Tom22, and Tom70. J. Biol. Chem. 272, 20730–20735 (1997). doi:10.1074/jbc.272.33.20730

14. J. Brix, G. A. Ziegler, K. Dietmeier, J. Schneider-Mergener, G. E. Schulz, N. Pfanner, The mitochondrial import receptor Tom70: Identification of a 25 kDa core domain with a specific binding site for preproteins. J. Mol. Biol. 303, 479–488 (2000). doi:10.1006/jmbi.2000.4120

15. R. D. Mills, J. Trewhella, T. W. Qiu, T. Welte, T. M. Ryan, T. Hanley, R. B. Knott, T. Lithgow, T. D. Mulhern, Domain organization of the monomeric form of the Tom70 mitochondrial import receptor. J. Mol. Biol. 388, 1043–1058 (2009). doi:10.1016/j.jmb.2009.03.070

16. S. D. Weeks, S. De Graef, A. Munawar, X-ray Crystallographic Structure of Orf9b from SARS-CoV-2 (2020); https://doi.org/10.2210/pdb6z4u/pdb.

17. M. Bouhaddou, D. Memon, B. Meyer, K. M. White, V. V. Rezelj, M. Correa Marrero, B. J. Polacco, J. E. Melnyk, S. Ulferts, R. M. Kaake, J. Batra, A. L. Richards, E. Stevenson, D. E. Gordon, A. Rojc, K. Obernier, J. M. Fabius, M. Soucheray, L. Miorin, E. Moreno, C. Koh, Q. D. Tran, A. Hardy, R. Robinot, T. Vallet, B. E. Nilsson-Payant, C. Hernandez- Armenta, A. Dunham, S. Weigang, J. Knerr, M. Modak, D. Quintero, Y. Zhou, A. Dugourd, A. Valdeolivas, T. Patil, Q. Li, R. Hüttenhain, M. Cakir, M. Muralidharan, M. Kim, G. Jang, B. Tutuncuoglu, J. Hiatt, J. Z. Guo, J. Xu, S. Bouhaddou, C. J. P. Mathy, A. Gaulton, E. J. Manners, E. Félix, Y. Shi, M. Goff, J. K. Lim, T. McBride, M. C. O'Neal, Y. Cai, J. C. J. Chang, D. J. Broadhurst, S. Klippsten, E. De Wit, A. R. Leach, T. Kortemme, B. Shoichet, M. Ott, J. Saez-Rodriguez, B. R. tenOever, R. D. Mullins, E. R. Fischer, G. Kochs, R. Grosse, A. García-Sastre, M. Vignuzzi, J. R. Johnson, K. M. Shokat, D. L. Swaney, P. Beltrao, N. J. Krogan, The Global Phosphorylation Landscape of SARS-CoV-2 Infection. Cell 182, 685–712.e19 (2020). doi:10.1016/j.cell.2020.06.034

18. J. Li, X. Qian, J. Hu, B. Sha, Molecular chaperone Hsp70/Hsp90 prepares the mitochondrial outer membrane translocon receptor Tom71 for preprotein loading. J. Biol. Chem. 284, 23852–23859 (2009). doi:10.1074/jbc.M109.023986

19. X.-Y. Liu, B. Wei, H.-X. Shi, Y.-F. Shan, C. Wang, Tom70 mediates activation of interferon regulatory factor 3 on mitochondria. Cell Res. 20, 994–1011 (2010). doi:10.1038/cr.2010.103

20. Y. Liu, C. Zhang, F. Huang, Y. Yang, F. Wang, J. Yuan, Z. Zhang, Y. Qin, X. Li, D. Zhao,

S. Li, S. Tan, Z. Wang, J. Li, C. Shen, J. Li, L. Peng, W. Wu, M. Cao, L. Xing, Z. Xu, L. Chen, C. Zhou, W. J. Liu, L. Liu, C. Jiang, Elevated plasma levels of selective cytokines in COVID-19 patients reflect viral load and lung injury. Natl. Sci. Rev. 7, 1003–1011 (2020). doi:10.1093/nsr/nwaa037

21. C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395, 497–506 (2020). doi:10.1016/S0140-6736(20)30183-5

22. C. Qin, L. Zhou, Z. Hu, S. Zhang, S. Yang, Y. Tao, C. Xie, K. Ma, K. Shang, W. Wang, D.-S. Tian, Dysregulation of Immune Response in Patients With Coronavirus 2019 (COVID-19) in Wuhan, China. Clin. Infect. Dis. 71, 762–768 (2020). doi:10.1093/cid/ciaa248

23. G. Chen, D. Wu, W. Guo, Y. Cao, D. Huang, H. Wang, T. Wang, X. Zhang, H. Chen, H. Yu, X. Zhang, M. Zhang, S. Wu, J. Song, T. Chen, M. Han, S. Li, X. Luo, J. Zhao, Q. Ning, Clinical and immunological features of severe and moderate coronavirus disease 2019. J. Clin. Invest. 130, 2620–2629 (2020). doi:10.1172/JCI137244

24. M. Zaretsky, R. Etzyoni, J. Kaye, L. Sklair-Tavron, A. Aharoni, Directed Evolution of a Soluble Human IL-17A Receptor for the Inhibition of Psoriasis Plaque Formation in a Mouse Model. Chem. Biol. 20, 202–211 (2013). doi:10.1016/j.chembiol.2012.11.012

25. M. Sohda, Y. Misumi, K. Tashiro, M. Yamazaki, T. Saku, K. Oda, Identification of a soluble isoform of human IL-17RA generated by alternative splicing. Cytokine 64, 642–645 (2013). doi:10.1016/j.cyto.2013.09.012

26. J. Lokau, C. Garbers, Biological functions and therapeutic opportunities of soluble cytokine receptors. Cytokine Growth Factor Rev. 55, 94–108 (2020). doi:10.1016/j.cytogfr.2020.04.003

27. M. Sammel, F. Peters, J. Lokau, F. Scharfenberg, L. Werny, S. Linder, C. Garbers, S. Rose- John, C. Becker-Pauly, Differences in Shedding of the Interleukin-11 Receptor by the Proteases ADAM9, ADAM10, ADAM17, Meprin α, Meprin β and MT1-MMP. Int. J. Mol. Sci. 20, 3677 (2019). doi:10.3390/ijms20153677

28. B. B. Sun, J. C. Maranville, J. E. Peters, D. Stacey, J. R. Staley, J. Blackshaw, S. Burgess, T. Jiang, E. Paige, P. Surendran, C. Oliver-Williams, M. A. Kamat, B. P. Prins, S. K. Wilcox, E. S. Zimmerman, A. Chi, N. Bansal, S. L. Spain, A. M. Wood, N. W. Morrell, J. R. Bradley, N. Janjic, D. J. Roberts, W. H. Ouwehand, J. A. Todd, N. Soranzo, K. Suhre, D. S. Paul, C. S. Fox, R. M. Plenge, J. Danesh, H. Runz, A. S. Butterworth, Genomic atlas of the human plasma proteome. Nature 558, 73–79 (2018). doi:10.1038/s41586-018-0175-2

29. Z. Zhu, Z. Zheng, F. Zhang, Y. Wu, M. Trzaskowski, R. Maier, M. R. Robinson, J. J. McGrath, P. M. Visscher, N. R. Wray, J. Yang, Causal associations between risk factors and common diseases inferred from GWAS summary data. Nat. Commun. 9, 224 (2018). doi:10.1038/s41467-017-02317-2

30. The COVID-19 Host Genetics Initiative, The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. Eur. J. Hum. Genet. 28, 715–718 (2020). doi:10.1038/s41431-020-0636-6

31. B. E. Young, S.-W. Fong, Y.-H. Chan, T.-M. Mak, L. W. Ang, D. E. Anderson, C. Y.-P. Lee, S. N. Amrun, B. Lee, Y. S. Goh, Y. C. F. Su, W. E. Wei, S. Kalimuddin, L. Y. A. Chai, S. Pada, S. Y. Tan, L. Sun, P. Parthasarathy, Y. Y. C. Chen, T. Barkham, R. T. P. Lin, S. Maurer-Stroh, Y.-S. Leo, L.-F. Wang, L. Renia, V. J. Lee, G. J. D. Smith, D. C. Lye, L. F. P. Ng, Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: An observational cohort study. Lancet 396, 603–611 (2020). doi:10.1016/S0140-6736(20)31757-8

32. C. Amici, A. Di Caro, A. Ciucci, L. Chiappa, C. Castilletti, V. Martella, N. Decaro, C. Buonavoglia, M. R. Capobianchi, M. G. Santoro, Indomethacin has a potent antiviral activity against SARS coronavirus. Antivir. Ther. 11, 1021–1030 (2006).

33. P. R. Rosenbaum, D. B. Rubin, The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55 (1983). doi:10.1093/biomet/70.1.41

34. C. Abate, P. D. Mosier, F. Berardi, R. A. Glennon, A structure-affinity and comparative molecular field analysis of sigma-2 (σ2) receptor ligands. Cent. Nerv. Syst. Agents Med. Chem. 9, 246–257 (2009). doi:10.2174/1871524910909030246

35. R. A. Glennon, Sigma receptor ligands and the use thereof, U.S. Patent 6,057,371 (2000); https://patentimages.storage.googleapis.com/dc/36/68/73f4ccdac4c973/US6057371.pdf.

36. R. R. Matsumoto, B. Pouw, Correlation between neuroleptic binding to σ1 and σ2 receptors and acute dystonic reactions. Eur. J. Pharmacol. 401, 155–160 (2000). doi:10.1016/S0014-2999(00)00430-1

37. M. Dold, M. T. Samara, C. Li, M. Tardy, S. Leucht, Haloperidol versus first-generation antipsychotics for the treatment of schizophrenia and other psychotic disorders. Cochrane Database Syst. Rev. 1, CD009831 (2015). doi:10.1002/14651858.CD009831.pub2

38. F. F. Moebius, R. J. Reiter, K. Bermoser, H. Glossmann, S. Y. Cho, Y. K. Paik, Pharmacological analysis of sterol delta8-delta7 isomerase proteins with [3H]ifenprodil. Mol. Pharmacol. 54, 591–598 (1998). doi:10.1124/mol.54.3.591

39. E. Gregori-Puigjané, V. Setola, J. Hert, B. A. Crews, J. J. Irwin, E. Lounkine, L. Marnett, B. L. Roth, B. K. Shoichet, Identifying mechanism-of-action targets for drugs and probes. Proc. Natl. Acad. Sci. U.S.A. 109, 11178–11183 (2012). doi:10.1073/pnas.1204524109

40. Z. Hubler, D. Allimuthu, I. Bederman, M. S. Elitt, M. Madhavan, K. C. Allan, H. E. Shick, E. Garrison, M. T Karl, D. C. Factor, Z. S. Nevin, J. L. Sax, M. A. Thompson, Y. Fedorov, J. Jin, W. K. Wilson, M. Giera, F. Bracher, R. H. Miller, P. J. Tesar, D. J. Adams, Accumulation of 8,9-unsaturated sterols drives oligodendrocyte formation and remyelination. Nature 560, 372–376 (2018). doi:10.1038/s41586-018-0360-3

41. F. F. Moebius, R. J. Reiter, M. Hanner, H. Glossmann, High affinity of sigma 1-binding sites for sterol isomerization inhibitors: Evidence for a pharmacological relationship with the yeast sterol C8-C7 isomerase. Br. J. Pharmacol. 121, 1–6 (1997). doi:10.1038/sj.bjp.0701079

42. H.-W. Jiang, H.-N. Zhang, Q.-F. Meng, J. Xie, Y. Li, H. Chen, Y.-X. Zheng, X.-N. Wang, H. Qi, J. Zhang, P.-H. Wang, Z.-G. Han, S.-C. Tao, SARS-CoV-2 Orf9b suppresses type I interferon responses by targeting TOM70. Cell. Mol. Immunol. 17, 998–1000 (2020). doi:10.1038/s41423-020-0514-8

43. J. J. Almagro Armenteros, C. K. Sønderby, S. K. Sønderby, H. Nielsen, O. Winther, DeepLoc: Prediction of protein subcellular localization using deep learning. Bioinformatics 33, 3387–3395 (2017). doi:10.1093/bioinformatics/btx431

44. C. Chiva, R. Olivella, E. Borràs, G. Espadas, O. Pastor, A. Solé, E. Sabidó, QCloud: A cloud-based quality control system for mass spectrometry-based proteomics laboratories. PLOS ONE 13, e0189209 (2018). doi:10.1371/journal.pone.0189209

45. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26, 1367–1372 (2008). doi:10.1038/nbt.1511

46. M. Giurgiu, J. Reinhard, B. Brauner, I. Dunger-Kaltenbach, G. Fobo, G. Frishman, C. Montrone, A. Ruepp, CORUM: The comprehensive resource of mammalian protein complexes-2019. Nucleic Acids Res. 47, D559–D563 (2019). doi:10.1093/nar/gky973

47. E. L. Huttlin, L. Ting, R. J. Bruckner, F. Gebreab, M. P. Gygi, J. Szpyt, S. Tam, G. Zarraga, G. Colby, K. Baltier, R. Dong, V. Guarani, L. P. Vaites, A. Ordureau, R. Rad, B. K. Erickson, M. Wühr, J. Chick, B. Zhai, D. Kolippakkam, J. Mintseris, R. A. Obar, T. Harris, S. Artavanis-Tsakonas, M. E. Sowa, P. De Camilli, J. A. Paulo, J. W. Harper, S. P. Gygi, The BioPlex Network: A Systematic Exploration of the Human Interactome. Cell 162, 425–440 (2015). doi:10.1016/j.cell.2015.06.043

48. G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, clusterProfiler: An R package for comparing biological themes among gene clusters. OMICS 16, 284–287 (2012). doi:10.1089/omi.2011.0118

49. M. Remmert, A. Biegert, A. Hauser, J. Söding, HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat. Methods 9, 173–175 (2011). doi:10.1038/nmeth.1818

50. J. Yang, I. Anishchenko, H. Park, Z. Peng, S. Ovchinnikov, D. Baker, Improved protein structure prediction using predicted interresidue orientations. Proc. Natl. Acad. Sci. U.S.A. 117, 1496–1503 (2020). doi:10.1073/pnas.1914677117

51. Y. Zhai, F. Sun, X. Li, H. Pang, X. Xu, M. Bartlam, Z. Rao, Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. Nat. Struct. Mol. Biol. 12, 980–986 (2005). doi:10.1038/nsmb999

52. A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296– W303 (2018). doi:10.1093/nar/gky427

53. J. Durairaj, M. Akdel, D. de Ridder, A. D. J. van Dijk, Geometricus Represents Protein Structures as Shape-mers Derived from Moment Invariants. bioRxiv 2020.09.07.285569 [Preprint]. 8 September 2020. https://doi.org/10.1101/2020.09.07.285569.

54. M. Akdel, J. Durairaj, D. de Ridder, A. D. J. van Dijk, Caretta - A multiple protein structure alignment and feature extraction suite. Comput. Struct. Biotechnol. J. 18, 981–992 (2020). doi:10.1016/j.csbj.2020.03.011

55. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504 (2003). doi:10.1101/gr.1239303

56. R. T. Pillich, J. Chen, V. Rynkov, D. Welker, D. Pratt, NDEx: A Community Resource for Sharing and Publishing of Biological Networks. Methods Mol. Biol. 1558, 271–301 (2017). doi:10.1007/978-1-4939-6783-4_13

57. D. K. W. Chu, Y. Pan, S. M. S. Cheng, K. P. Y. Hui, P. Krishnan, Y. Liu, D. Y. M. Ng, C. K. C. Wan, P. Yang, Q. Wang, M. Peiris, L. L. M. Poon, Molecular Diagnosis of a Novel Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia. Clin. Chem. 66, 549–555 (2020). doi:10.1093/clinchem/hvaa029

58. K. J. Livak, T. D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402–408 (2001). doi:10.1006/meth.2001.1262

59. R. Stoner, T. Maures, D. Conant, Methods and systems for guide RNA design and use, U.S. Patent 2019/0382797 A1 (2019); https://patentimages.storage.googleapis.com/95/c7/43/3d48387ce0f116/US20190382797A1.pdf

60. T. Hsiau, D. Conant, N. Rossi, T. Maures, K. Waite, J. Yang, S. Joshi, R. Kelso, K. Holden, B. L. Enzmann, R. Stoner, Inference of CRISPR Edits from Sanger Trace Data. bioRxiv 251082 [Preprint]. 10 August 2018. https://doi.org/10.1101/251082.

61. A. S. Jureka, J. A. Silvas, C. F. Basler, Propagation, Inactivation, and Safety Testing of SARS-CoV-2. Viruses 12, 622 (2020). doi:10.3390/v12060622

62. A. C. Y. Fan, M. K. Bhangoo, J. C. Young, Hsp90 functions in the targeting and outer membrane translocation steps of Tom70-mediated mitochondrial import. J. Biol. Chem. 281, 33313–33324 (2006). doi:10.1074/jbc.M605250200

63. S. Backes, S. Hess, F. Boos, M. W. Woellhaf, S. Gödel, M. Jung, T. Mühlhaus, J. M. Herrmann, Tom70 enhances mitochondrial preprotein import efficiency by binding to internal targeting sequences. J. Cell Biol. 217, 1369–1382 (2018). doi:10.1083/jcb.201708044

64. J. J. Almagro Armenteros, M. Salvatore, O. Emanuelsson, O. Winther, G. von Heijne, A. Elofsson, H. Nielsen, Detecting sequence signals in targeting peptides using deep learning. Life Sci. Alliance 2, e201900429 (2019). doi:10.26508/lsa.201900429

65. Y. Fukasawa, J. Tsuji, S.-C. Fu, K. Tomii, P. Horton, K. Imai, MitoFates: Improved prediction of mitochondrial targeting sequences and their cleavage sites. Mol. Cell. Proteomics 14, 1113–1126 (2015). doi:10.1074/mcp.M114.043083

66. A. Drozdetskiy, C. Cole, J. Procter, G. J. Barton, JPred4: A protein secondary structure prediction server. Nucleic Acids Res. 43, W389–W394 (2015). doi:10.1093/nar/gkv332

67. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. J. Struct. Biol. 152, 36–51 (2005). doi:10.1016/j.jsb.2005.07.007

68. S. Q. Zheng, E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng, D. A. Agard, MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods 14, 331–332 (2017). doi:10.1038/nmeth.4193

69. S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402 (1997). doi:10.1093/nar/25.17.3389

70. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. Acta Cryst. D60, 2126–2132 (2004). doi:10.1107/S0907444904019158

71. R. Y.-R. Wang, Y. Song, B. A. Barad, Y. Cheng, J. S. Fraser, F. DiMaio, Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. eLife 5, e17219 (2016). doi:10.7554/eLife.17219

72. R. T. Kidmose, J. Juhl, P. Nissen, T. Boesen, J. L. Karlsen, B. P. Pedersen, Namdinator - automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. IUCrJ 6, 526–531 (2019). doi:10.1107/S2052252519007619

73. T. I. Croll, ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. Acta Cryst. D74, 519–530 (2018). doi:10.1107/S2059798318002425

74. T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris, T. E. Ferrin, UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci. 27, 14–25 (2018). doi:10.1002/pro.3235

75. P. V. Afonine, B. P. Klaholz, N. W. Moriarty, B. K. Poon, O. V. Sobolev, T. C. Terwilliger, P. D. Adams, A. Urzhumtsev, New tools for the analysis and validation of cryo-EM maps and atomic models. Acta Cryst. D74, 814–840 (2018). doi:10.1107/S2059798318009324

76. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797 (2007). doi:10.1016/j.jmb.2007.05.022

77. A. N. Honko, N. Storm, D. J. Bean, J. H. Vasquez, S. N. Downs, A. Griffiths, Rapid Quantification and Neutralization Assays for Novel Coronavirus SARS-CoV-2 Using Avicel RC 591 Semi-Solid Overlay. Preprints 2020050264 [Preprint]. 16 May 2020. www.preprints.org/manuscript/202005.0264/v1.

78. A. Šali, T. L. Blundell, Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779–815 (1993). doi:10.1006/jmbi.1993.1626

79. T. Yamada, J. Komoto, K. Watanabe, Y. Ohmiya, F. Takusagawa, Crystal structure and possible catalytic mechanism of microsomal prostaglandin E synthase type 2 (mPGES-2). J. Mol. Biol. 348, 1163–1176 (2005). doi:10.1016/j.jmb.2005.03.035

80. W. Yin, C. Mao, X. Luan, D.-D. Shen, Q. Shen, H. Su, X. Wang, F. Zhou, W. Zhao, M. Gao, S. Chang, Y.-C. Xie, G. Tian, H.-W. Jiang, S.-C. Tao, J. Shen, Y. Jiang, H. Jiang, Y. Xu, S. Zhang, Y. Zhang, H. E. Xu, Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. Science 368, 1499–1504 (2020). doi:10.1126/science.abc1560

81. D. Kozakov, D. R. Hall, B. Xia, K. A. Porter, D. Padhorny, C. Yueh, D. Beglov, S. Vajda, The ClusPro web server for protein-protein docking. Nat. Protoc. 12, 255–278 (2017). doi:10.1038/nprot.2016.169

82. B. G. Pierce, K. Wiehe, H. Hwang, B.-H. Kim, T. Vreven, Z. Weng, ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers. Bioinformatics 30, 1771–1773 (2014). doi:10.1093/bioinformatics/btu097

83. Y. Yan, H. Tao, J. He, S.-Y. Huang, The HDOCK server for integrated protein-protein docking. Nat. Protoc. 15, 1829–1852 (2020). doi:10.1038/s41596-020-0312-x

84. A. Tovchigrechko, I. A. Vakser, GRAMM-X public web server for protein-protein docking. Nucleic Acids Res. 34, W310–W314 (2006). doi:10.1093/nar/gkl206

85. M. Torchala, I. H. Moal, R. A. G. Chaleil, J. Fernandez-Recio, P. A. Bates, SwarmDock: A server for flexible protein-protein docking. Bioinformatics 29, 807–809 (2013). doi:10.1093/bioinformatics/btt038

86. D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, H. J. Wolfson, PatchDock and SymmDock: Servers for rigid and symmetric docking. Nucleic Acids Res. 33, W363–W367 (2005). doi:10.1093/nar/gki481

87. G. Q. Dong, H. Fan, D. Schneidman-Duhovny, B. Webb, A. Sali, Optimized atomic statistical potentials: Assessment of protein interfaces and loops. Bioinformatics 29, 3158–3166 (2013). doi:10.1093/bioinformatics/btt560

88. J. Armstrong, G. Hickey, M. Diekhans, A. Deran, Q. Fang, D. Xie, S. Feng, J. Stiller, D. Genereux, J. Johnson, V. D. Marinescu, D. Haussler, J. Alföldi, K. Lindblad-Toh, E. Karlsson, E. D. Jarvis, G. Zhang, B. Paten, Progressive alignment with Cactus: A multiple-genome aligner for the thousand-genome era. bioRxiv 730531 [Preprint]. 15 October 2019. https://doi.org/10.1101/730531.

89. B. Paten, D. Earl, N. Nguyen, M. Diekhans, D. Zerbino, D. Haussler, Cactus: Algorithms for genome multiple sequence alignment. Genome Res. 21, 1512–1528 (2011). doi:10.1101/gr.123356.111

90. M. D. Smith, J. O. Wertheim, S. Weaver, B. Murrell, K. Scheffler, S. L. Kosakovsky Pond, Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol. Biol. Evol. 32, 1342–1353 (2015). doi:10.1093/molbev/msv022

91. S. L. K. Pond, S. D. W. Frost, S. V. Muse, HyPhy: Hypothesis testing using phylogenies. Bioinformatics 21, 676–679 (2005). doi:10.1093/bioinformatics/bti079

92. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 20, 110–121 (2010). doi:10.1101/gr.097857.109

93. M. J. Hubisz, K. S. Pollard, A. Siepel, PHAST and RPHAST: Phylogenetic analysis with space/time models. Brief. Bioinform. 12, 41–51 (2011). doi:10.1093/bib/bbq072

94. R. Ramani, K. Krumholz, Y.-F. Huang, A. Siepel, PhastWeb: A web interface for evolutionary conservation scoring of multiple sequence alignments using phastCons and phyloP. Bioinformatics 35, 2320–2322 (2019). doi:10.1093/bioinformatics/bty966

95. W. A. Ray, Evaluating medication effects outside of clinical trials: New-user designs. Am. J. Epidemiol. 158, 915–920 (2003). doi:10.1093/aje/kwg231

96. S. Schneeweiss, A basic study design for expedited safety signal evaluation based on electronic healthcare data. Pharmacoepidemiol. Drug Saf. 19, 858–868 (2010). doi:10.1002/pds.1926

97. H. Quan, V. Sundararajan, P. Halfon, A. Fong, B. Burnand, J.-C. Luthi, L. D. Saunders, C.A. Beck, T. E. Feasby, W. A. Ghali, Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med. Care 43, 1130–1139 (2005). doi:10.1097/01.mlr.0000182534.19832.83

98. P. C. Austin, Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat. Med. 28, 3083– 3107 (2009). doi:10.1002/sim.3697

99. World Health Organization (WHO), WHO R&D Blueprint, novel Coronavirus: COVID-19 Therapeutic Trial Synopsis (WHO, 2020); www.who.int/blueprint/priority-diseases/key-action/COVID- 19_Treatment_Trial_Design_Master_Protocol_synopsis_Final_18022020.pdf.

100. Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D. J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Pérez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, S. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A. F. Jarnuczak, T. Ternent, A. Brazma, J. A. Vizcaíno, The PRIDE database and related tools and resources in 2019: Improving support for quantification data. Nucleic Acids Res. 47, D442–D450 (2019). doi:10.1093/nar/gky1106

101. J. Li, X. Qian, J. Hu, B. Sha, Crystal structure of Tom71 complexed with Hsp82 C-terminal fragment (2009); https://doi.org/10.2210/pdb3fp2/pdb

Chapter 4

The Cellular NMD Pathway Restricts Zika Virus Infection and is Targeted by the Viral Capsid Protein

Krystal A. Fontaine,# Kristoffer E. Leon,# **Mir M. Khalid,** Sakshi Tomar, David Jimenez-Morales, Mariah Dunlap, Julia A. Kaye, Priya S. Shah, Steve Finkbeiner, Nevan J. Krogan, Melanie Ott.*

mBio, Volume 9 Issue 6 e02126-18

DOI: 10.1128/mBio.02126-18

#These authors contributed equally *Lead contact

The cellular NMD pathway restricts Zika virus infection and is targeted by the viral capsid protein

Krystal A. Fontaine[a], Kristoffer E. Leon[a,b], Mir M. Khalida, Sakshi Tomara, David Jimenez-Morales[a,c,*], Mariah Dunlap[a], Julia A. Kaye[a], Priya S. Shah[c,*], Steve Finkbeiner[a,d], Nevan J. Krogan[a,c], Melanie Ott[a,e,#].

[a]Gladstone Institutes, San Francisco, California, USA
[b]Medical Scientist Training Program and Biomedical Sciences Graduate Program, University of California, San Francisco, California, USA
[c]Quantitative Biosciences Institute (QBI) and Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California, USA
[d]Departments of Neurology and Physiology, University of California, San Francisco, California, USA
[e]Department of Medicine, University of California, San Francisco, California, USA
[#]Address correspondence to Melanie Ott, melanie.ott@gladstone.ucsf.edu.

*Present address: Priya S. Shah, Departments of Chemical Engineering and Microbiology and Molecular Genetics, University of California, Davis, CA. David Jimenez-Morales, PhD, Department of Medicine, Division of Cardiovascular Medicine, Stanford University.

K.A.F and K.E.L contributed equally to this work.

ABSTRACT

Zika virus (ZIKV) infection of neural progenitor cells (NPCs) in utero is associated with neurological disorders, such as microcephaly, but a detailed molecular understanding of ZIKV-induced pathogenesis is lacking. Here we show that in vitro ZIKV infection of human cells, including NPCs, causes disruption of the nonsense-mediated mRNA decay (NMD) pathway. NMD is a cellular mRNA surveillance mechanism that is required for normal brain size in mice. Using affinity purification-mass spectrometry, we identified multiple cellular NMD factors that bind to the viral capsid protein, including the central NMD regulator up-frameshift protein 1 (UPF1). Endogenous UPF1 interacted with the ZIKV capsid protein in coimmunoprecipitation experiments, and capsid expression posttranscriptionally downregulated UPF1 protein levels, a process that we confirmed occurs during ZIKV infection. Cellular fractionation studies show that the ZIKV capsid protein specifically targets nuclear UPF1 for degradation via the proteasome. A further decrease in UPF1 levels by RNAi significantly enhanced ZIKV infection in NPC cultures, consistent with a model in which NMD restricts ZIKV infection in the fetal brain. We propose that ZIKV, via the capsid protein, has evolved a strategy to lower UPF1 levels and dampen antiviral activities of NMD, which in turn contributes to neuropathology in vivo.

IMPORTANCE

Zika virus (ZIKV) is a significant global health threat, as infection has been linked to serious neurological complications, including microcephaly. Using a human stem cell-derived neural progenitor model system, we find that a critical cellular quality control process called the nonsense-mediated mRNA decay (NMD) pathway is disrupted during ZIKV infection. Importantly, disruption of the NMD pathway is a known cause of microcephaly and other neurological disorders. We further identify an interaction between the capsid protein of ZIKV and up-frameshift protein 1 (UPF1), the master regulator of NMD, and show that ZIKV capsid targets UPF1 for degradation. Together, these results offer a new mechanism for how ZIKV infection can cause neuropathology in the developing brain.

KEYWORDS Zika virus, nonsense-mediated mRNA decay pathway, virus-host interactions.

INTRODUCTION

Zika virus (ZIKV) is a mosquito-borne RNA virus that belongs to the Flaviviridae family. First isolated in Uganda in 1947, ZIKV remained relatively obscure for decades following its discovery because infection was associated with only mild disease. However, more severe clinical manifestations, including microcephaly, have been observed during the recent spread of ZIKV through the Americas (1). ZIKV infection induces cell cycle arrest and apoptosis in neural progenitor cells (NPCs) in in vitro studies and in vivo mouse models, with the latter resulting in cortical thinning and microcephaly (2–6). While it is now established that ZIKV infection during pregnancy is a causative agent of microcephaly (7), the molecular mechanisms underlying ZIKV-induced neuropathogenesis remain largely unknown.

Similar to other flaviviruses, ZIKV contains a single-stranded, positive-sense RNA genome of ~11 kb in size. The genome encodes a single polyprotein that is posttranslationally processed by both host and viral proteases to produce 3 structural proteins and 7 nonstructural proteins (8, 9). The flavivirus capsid, which is the first protein encoded in the genome, is a major structural element required for the encapsidation of the RNA genome during virion assembly (10). While flavivirus replication is known to occur in the cytoplasm, a significant portion of the viral capsid protein localizes to the nucleus during infection (10, 11). Although the role of nuclear capsid during infection is less clear, several functions have been suggested. The capsid protein from dengue virus, a close relative of ZIKV, binds to core histones and inhibits nucleosome formation, thus implicating the protein in altering host gene expression (12). Furthermore, several flavivirus capsid proteins, including ZIKV capsid, localize to the nucleolus, with many interacting with nucleolar proteins to promote viral particle production (13–16).

The nonsense-mediated mRNA decay (NMD) pathway was initially discovered as a highly conserved quality control system that destroys transcripts containing premature termination codons (PTCs) (17). Following splicing of pre-mRNAs, a multisubunit protein complex called the exon-junction complex (EJC) is deposited onto mRNAs near the sites of exon-exon junctions. If a PTC is found ~50 to 55 nucleotides upstream of an EJC, the mRNA will be subjected to NMD-mediated degradation initiated by the recruitment of the RNA helicase up-frameshift protein 1 (UPF1). UPF1 plays a central role in the NMD pathway by linking the translation termination event to the assembly of a surveillance complex, resulting in NMD activation (18). Interestingly, microcephaly has been associated with genetic mutations that result in the impairment of the NMD pathway. While knockout of Upf1 and other NMD factors is embryonic lethal in mice (19),

mice haploinsufficient for the EJC components Magoh, Rbm8a, and Eif4a3 exhibit aberrant neurogenesis and microcephaly (20–22).

In addition to PTC-containing transcripts, it is now known that the NMD pathway recognizes a broader range of RNA substrates. Notably, the NMD controls the “normal” expression of ~10% of the cellular transcriptome and is regarded as a posttranscriptional mechanism of gene regulation (23). Furthermore, the NMD pathway also regulates viral infections. While it was first reported that UPF1 promotes the infectivity of HIV-1 progeny virions (24), replication of several human RNA viruses, including human T-cell lymphotropic virus type 1 (HTLV-1), Semliki Forest virus, and Sindbis virus, is enhanced following UPF1 knockdown, implicating UPF1 and the NMD pathway, either directly or indirectly, in the host antiviral response (25–28). As ZIKV infection and NMD impairment both promote microcephaly development, and we previously described disruption of the NMD pathway in cells infected with a related flavivirus, the hepatitis C virus (HCV) (29), we hypothesized that ZIKV infection manipulates the cellular NMD pathway, a process contributing to ZIKV-induced neuropathology.

RESULTS

The NMD pathway is impaired during ZIKV infection.

To determine whether ZIKV infection affects NMD, we infected human hepatoma cells (Huh7) and human induced pluripotent stem cell (iPSC)-derived NPCs with ZIKV for 48 h. We isolated total RNA from infected cells and measured mRNA levels of three canonical NMD substrates: asparagine synthetase (ASNS), cysteinyl-tRNA synthetase (CARS), and SR protein SC35 (29). ASNS, CARS, and SC35 transcripts were significantly elevated in Huh7 cells and NPCs following infection with Asian lineage ZIKV strain P6-740 (Fig. 1a). Levels of NMD substrates were also elevated in Huh7 cells infected with the contemporary ZIKV clinical isolate PRVABC59 (Puerto Rico, 2015) (Fig. 1a). We found that the ZIKV-induced increase in NMD transcripts did not reflect a global increase in transcription, as mRNA levels of housekeeping genes, including glyceraldehyde 3-phosphate dehydrogenase (GAPDH), were not altered in infected cells (Fig. 1a). Together, these results indicate that the NMD pathway is impaired in ZIKV-infected cells.

NMD substrates are regulated through the ATP-dependent RNA helicase activity of UPF1, the central regulator of NMD (18). To determine whether ZIKV infection broadly affects NMD, we utilized two publicly available RNA sequencing (RNA-Seq) data sets to compare genome-wide transcriptional alterations found during ZIKV infection (6) to those found following UPF1 knockdown (30). As shown in Fig. 1b, there is a significant overlap in upregulated genes between these two data sets. Interestingly, several of the overlapping genes are canonical NMD substrates (31–35) involved in cell cycle arrest and induction of apoptosis, two conditions linked to ZIKV-associated neuropathology (5). These genes include DNA damage-inducible transcript 3 (DDIT3) (36) and growth arrest and DNA damage-inducible protein 45 alpha and beta (GADD45A and GADD45B, respectively) (37). Via quantitative real-time reverse transcription-PCR (qRT-PCR), we confirmed that transcripts of each were upregulated following infection of Huh7 cells with ZIKV PRVABC59, while the mRNA levels of the housekeeping genes GAPDH, hypoxanthine phosphoribosyltransferase 1 (HPRT1), and lactate dehydrogenase A (LDHA) were not elevated (Fig. 1c). Combined, these data show that ZIKV infection is associated with dysregulated expression of NMD substrates relevant to ZIKV-mediated neuropathogenesis.

ZIKV capsid interacts with the NMD pathway.

We previously showed that the core protein of HCV, as well as the capsid protein of the related flaviviruses dengue virus and West Nile virus, interact with the host protein within bgcn homolog (WIBG/PYM1), an EJC disassembly factor associated with NMD (29, 38). To examine potential

interactions between ZIKV and the NMD pathway, we separately analyzed data generated from an affinity purification-mass spectrometry (AP-MS) screen to specifically query whether the capsid protein of ZIKV interacts with NMD-associated host factors (P. S. Shah et al., submitted for publication). ZIKV-host protein-protein interaction (PPI) maps were generated in HEK293T cells using ZIKV proteins from the Ugandan 1947 strain MR 766 or the French Polynesian 2013 strain H/PF/2013 as bait proteins.

From this analysis, we found that ZIKV capsid proteins interacted with several components of the NMD pathway, including multiple members of the EJC complex, as well as UPF1 and UPF3B, the latter an NMD effector that stimulates UPF1 helicase activity (Fig. 2a) (39). Importantly, the NMD host factors that interacted with each of the two different capsid proteins greatly overlapped, revealing that the interaction between capsid and the NMD pathway is conserved across the Asian and African lineages of ZIKV (Fig. 2a). Next, we validated the binding of ZIKV capsid to select NMD host factors by coimmunoprecipitating FLAG-tagged capsid protein with endogenous UPF3B or UPF1 in HEK293T cells. Both UPF3B and UPF1 proteins coimmunoprecipitated with ZIKV capsid, thus confirming the AP-MS results (Fig. 2b and c, respectively). UPF1 interacted with the viral capsid protein independently from its RNA-binding and ATPase/helicase capacities, as ZIKV capsid coimmunoprecipitated with UPF1 mutants deficient in these functions following overexpression (Fig. 2d).

ZIKV capsid degrades UPF1, the master regulator of NMD.

Surprisingly, we consistently observed a decrease in UPF1, but not UPF3B, protein levels in the input lysate of ZIKV capsid-transfected cells, pointing to a specific perturbation of endogenous UPF1 expression by ZIKV capsid (Fig. 2c). To confirm that UPF1 protein levels are dysregulated during ZIKV infection, we performed Western blot analysis of infected Huh7 cells and NPCs. Cellular UPF1 protein levels were consistently downregulated by ~50% in ZIKV-infected Huh7 cells, whereas an ~25% reduction was observed in ZIKV-infected NPCs (Fig. 2e), mirroring the difference in infection efficiencies achieved in these two cell systems. UPF1 transcript levels were not decreased in ZIKV-infected cells or following ZIKV capsid overexpression, indicating that UPF1 is posttranscriptionally downregulated during ZIKV infection (see Fig. S1a and b, respectively, in the supplemental material).

Because ZIKV capsid and UPF1 both localize to the nucleus and the cytoplasm (13, 40), we performed fractionation studies in ZIKV capsid-transfected HEK293T cells to determine whether UPF1 is downregulated within a specific cellular compartment. Capsid expression markedly

decreased nuclear UPF1 levels, whereas cytoplasmic levels were unchanged (Fig. 3a). We next examined a potential role for the autophagic and proteasomal pathways, both of which are known to mediate nuclear protein degradation, in ZIKV capsid-induced UPF1 downregulation. As shown in Fig. S2, nuclear UPF1 levels in ZIKV capsid-transfected cells were not rescued by inhibition of cellular autophagy via bafilomycin A1 treatment. However, nuclear UPF1 levels were restored in a dose-dependent manner following treatment with the proteasome inhibitor bortezomib (Fig. 3b), indicating enhanced proteasomal degradation of nuclear UPF1 in the presence of ZIKV capsid. Although ZIKV capsid colocalized with endogenous UPF1 in the cytoplasm of transfected Huh7-Lunet cells (Mander's colocalization coefficient of ~57%), we detected very little colocalization within the nucleus (~7%) (Fig. S3). We hypothesized that this was due to specific degradation of UPF1 by ZIKV capsid in the nucleus. Indeed, when cells were treated with bortezomib, the fraction of nuclear UPF1 colocalizing with ZIKV capsid increased ~8-fold, while the fraction of nuclear capsid interacting with UPF1 remained unchanged (Fig. 3c). These results demonstrate that ZIKV capsid interacts with UPF1 both in the cytoplasm and nucleus but that it specifically targets nuclear UPF1 for proteasomal degradation.

UPF1 is a restriction factor of ZIKV.

To test the effect of lowered UPF1 levels on ZIKV infection, we further decreased UPF1 expression prior to ZIKV infection by transfecting NPCs with either nontargeting siRNA or a pool of UPF1-specific siRNAs. We then infected the transfected cells with ZIKV and measured viral RNA levels, as well as infectious titers, 48 h postinfection (hpi). UPF1 knockdown was successful in siRNA-treated cells, as confirmed by Western blot analysis (Fig. 4a). The depletion of UPF1 in NPCs prior to infection resulted in a significant increase in both ZIKV RNA levels and infectious virus production (Fig. 4b and c, respectively), indicating that expression of UPF1 restricts ZIKV infection at or before the RNA replication stage. To differentiate between these two stages, we analyzed double-stranded RNA (dsRNA) intermediates representing presumed viral RNA replication centers in infected NPCs (Fig. 4d) (41). Using confocal microscopy and 3D reconstruction analyses, we observed no significant difference in the number and size of dsRNA foci per cell when comparing ZIKV-infected, UPF1-depleted NPCs to ZIKV-infected cultures expressing UPF1 (Fig. 4e and f). Instead, we found a significant increase in the number of infected cells in NPC cultures when UPF1 was depleted, indicating that UPF1 regulates permissivity of NPCs to ZIKV infection at an early stage prior to viral RNA replication (Fig. 4g).

DISCUSSION

In summary, we identified the NMD pathway as a restriction mechanism for ZIKV infection in human NPCs. NMD was partially inactivated in ZIKV-infected NPCs through expression of the viral capsid protein and the resulting degradation of host nuclear UPF1. As further weakening NMD by depleting UPF1 resulted in a marked increase in the number of infected cells, we propose a model in which an evolutionary “arms race” between cellular NMD and ZIKV determines whether a cell is successfully infected (Fig. 4h).

Downregulation of UPF1 by ZIKV capsid is not complete and is likely limited by the damaging effects of NMD disruption, as illustrated by the upregulation of genes involved in cell growth arrest and apoptosis. Indeed, knockout of Upf1 and other members of the NMD pathway is embryonic lethal in mice (19). However, mice haploinsufficient for NMD factors upstream of UPF1, including Magoh, Rbm8a, and Eif4a3, develop microcephaly (20–22). Thus, the reduction in nuclear UPF1 we observe in ZIKV-infected NPCs could contribute to the microcephaly phenotype caused by ZIKV infection in the fetal brain. While fetal and adult NPCs appear to be transcriptionally distinct (42), it has been shown that adult NPCs are also permissive to ZIKV infection (43). As the NMD pathway is a ubiquitous cellular surveillance mechanism, it is likely that ZIKV capsid targets UPF1 for degradation in any cell type that is susceptible to ZIKV infection. Accordingly, we have found that UPF1 is degraded following infection of both NPCs and hepatic Huh7 cells.

Why ZIKV capsid specifically downregulates nuclear UPF1 and how nuclear UPF1 contributes to ZIKV restriction remain unanswered. Several studies suggest that NMD is associated with the nucleus, although this issue remains controversial. Multiple transcripts, such as those encoding T cell receptor beta, triosephosphate isomerase, and mouse major urinary protein, have been shown to be specifically degraded in purified nuclei or reduced in nuclear fractions (44). These data support the model that selectively depleting nuclear UPF1 levels disrupts NMD function in ZIKV-infected cells. In addition, UPF1 is involved in several other processes within the nucleus, including nucleus-associated RNA metabolism, cell cycle progression, and DNA replication (40). Therefore, by targeting nuclear UPF1, ZIKV could disrupt these processes and program target cells for viral replication. Notably, viral RNA replication is thought to occur solely within the cytoplasmic compartment (45, 46). Using confocal microscopy and 3D reconstruction, we did not detect dsRNA foci localized within the nuclei of ZIKV-infected cells (data not shown), supporting our finding that UPF1 does not restrict viral RNA replication. While our results suggest a role for nuclear UPF1 in ZIKV restriction, it is possible that UPF1 also serves as a restriction factor of ZIKV within the cytoplasm. Previously, it was shown that UPF1 suppresses alphavirus replication

by degrading the incoming viral RNA following uncoating in the cytosol (28). Thus, ZIKV may possess an additional mechanism to prevent cytoplasmic UPF1 from targeting its incoming RNA genome for destruction.

Our data reveal that nuclear UPF1 is degraded by ZIKV capsid in a proteasome-dependent manner. While the nuclear proteasome has not been specifically linked to microcephaly, it plays critical roles in the regulation of chromatin structure, gene expression, DNA repair, and protein quality control (47). Thus, the co-opting of the nuclear proteasome by ZIKV capsid to degrade UPF1 could disrupt its normal proteasomal activity and further contribute to the cytopathic effects associated with ZIKV infection. Furthermore, given that the capsid protein of the closely related dengue virus can translocate across cell membranes (48), it is possible that capsid released from apoptotic, ZIKV-infected cells can enter neighboring, uninfected cells to degrade UPF1, thus increasing permissivity of bystander cells to ZIKV infection. Studies are ongoing to determine the precise molecular mechanism of ZIKV capsid-mediated UPF1 degradation and how UPF1 depletion enhances ZIKV replication, directly or indirectly. Ultimately, these data may help inform new therapeutic approaches, as reinforcement of the antiviral properties of the NMD pathway is expected to enhance resistance of NPCs to ZIKV infection and to promote normal neurodevelopment in infected fetuses.

FIG 1: ZIKV infection disrupts the NMD pathway. (a) Transcript levels of NMD substrates and a housekeeping gene control from Huh7 cells or NPCs mock infected or infected with ZIKV strain P6-740 or the contemporary clinical isolate PRVABC59. Cells were infected at a multiplicity of infection (MOI) of 0.1 or 1 and harvested at 48 hours postinfection (hpi). Data are represented as means plus standard errors of the means (SEM) (error bars). Statistical significance (and P values) were calculated by unpaired Student's t test and indicated as follows: *, $P \leq 0.05$; **, $P \leq 0.01$; ns, not significant. Three independent experiments were performed. (b) Venn diagram showing overlap of significantly upregulated genes associated with ZIKV infection of NPCs and UPF1 knockdown (KD) in HeLa cells. RNA-Seq analyses of mock-infected or ZIKV-infected NPCs harvested at 56 hpi and control siRNA-treated or UPF1 siRNA-treated HeLa TO cells harvested at 72 h posttransfection (hpt). The GeneProf hypergeometric probability calculator was then used to generate a hypergeometric P value. ****, $P \leq 0.0001$. (c) Transcript levels of housekeeping genes and select genes involved in cell cycle arrest and apoptosis that were identified in panel b. Huh7 cells were mock infected or infected with ZIKV PRVABC59 at an MOI of 0.1 or 1 and harvested at 48 hpi. Data are represented as means plus SEM. The P values were calculated by unpaired Student's t test. *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ns, not significant. n = 3 independent experiments.

FIG 2: The capsid protein of ZIKV interacts with the NMD pathway. (a) Ugandan ZIKV capsid (Ug Cap) (MR 766) and French Polynesian ZIKV capsid (Fp Cap) (H/PF/2013) PPI maps that show significant enrichment for host NMD-associated factors (purple), as identified by AP-MS (SAINTq probability score of > 0.9 and false-discovery rate [FDR] of <0.05). Ten interactions between Fp Cap and host NMD factors (hypergeometrical test, P value = $7.16 \times 10{-}10$) and eight interactions between Ug Cap and host NMD factors (P value = $3.45 \times 10{-}7$) were identified. (b) Coimmunoprecipitation (co-IP) and Western blot analysis of HEK293T cells transfected with vector or Flag-tagged ZIKV capsid (H/PF/2013, Asian lineage) and harvested at 48 hpt to immunoprecipitate endogenous UPF3B. The upper band detected in the IP Capsid blot represents a nonspecific artifact. α-Flag, anti-Flag antibody. (c) Co-IP and Western blot analysis of HEK293T cells transfected with vector or Flag-tagged ZIKV capsid and harvested at 48 hpt to immunoprecipitate endogenous UPF1. (d) Myc tag co-IP and Western blot analysis of HEK293T cells transfected with Strep-tagged ZIKV capsid and Myc-UPF1 (wild type), Myc-UPF1-C126S (RNA-binding mutant) or Myc-UPF1-G495R/G497E (ATPase/helicase mutant) and harvested at 48 hpt to immunoprecipitate ZIKV capsid. (e) Western blot analysis of UPF1 levels in mock-infected and ZIKV-infected (PRVABC59, MOI of 1) Huh7 cells or mock-infected and ZIKV-infected (P6-740, MOI of 1) NPCs harvested at 48 hpi, with β-actin and ZIKV envelope (ZIKV E) or ZIKV

capsid (ZIKV C) protein serving as loading and infection controls, respectively. Densitometric analyses were performed using ImageJ to quantify relative band intensities. Data are represented as means plus SEM. The P values were calculated by unpaired Student's t test. **, $P \leq 0.01$; ***, $P \leq 0.001$. n = 3 independent experiments.

FIG 3: ZIKV capsid degrades UPF1, the master regulator of NMD, via a proteasome-dependent mechanism. (a) Western blot analysis of UPF1 levels in subcellular fractionated HEK293T cells transfected with vector or Flag-tagged ZIKV capsid (H/PF/2013, Asian lineage) for 48 h. GAPDH was used as a cytoplasmic marker, and SP1 was used as a nuclear marker to ensure optimal fractionation. Densitometric analyses were performed using ImageJ to quantify relative band intensities. Data are represented as means plus SEM. P values were calculated by unpaired Student's t test. **, $P \leq 0.01$; ns, not significant. n = 3 independent experiments. (b) Western blot analysis of nuclear UPF1 levels in fractionated HEK293T cells transfected with vector or Flag-tagged ZIKV capsid for 48 h. Cells were treated with DMSO or increasing concentrations of the proteasome inhibitor bortezomib (Borte) for 24 h before harvest. Densitometric analyses were performed using ImageJ to quantify relative band intensities. Data are represented as means plus SEM. P values were calculated by one-way ANOVA with multiple comparisons. *, $P \leq 0.05$; ns, not significant. n = 3 independent experiments. (c) Representative 3D confocal microscopy images of the nuclei of Huh7-Lunet cells transfected with Strep-tagged ZIKV capsid. Cells were treated at 24 hpt with DMSO or 10 nM bortezomib and processed for immunostaining at 48 hpt with antibodies against Strep tag (turquoise) and endogenous UPF1 (purple). DAPI (blue) was used to stain and define the nuclei. Each channel was reconstructed digitally for visualization of the 3D colocalization. The thresholded Mander's correlation coefficients were determined, and P values were calculated by unpaired Student's t test. **, $P \leq 0.01$. n = 8 cells per condition. Scale bars represent 3 µm.

FIG 4: UPF1 knockdown enhances the permissivity of NPCs to ZIKV infection. (a) Western blot analysis of UPF1 levels in NPCs transfected with nontargeting siRNA (siNT) or a pool of UPF1-specific siRNAs (siUPF1) at 96 hpt. Densitometric analyses were performed using ImageJ to quantify relative band intensities. Data are represented as means plus SEM. P values were calculated by unpaired Student's t test. **, $P \leq 0.01$. n = 3 independent experiments using one NPC line. (b) ZIKV RNA levels in siNT-treated or siUPF1-treated NPCs infected with ZIKV strain PRVABC59 at an MOI of 0.1 or 1 and harvested at 48 hpi. Data are represented as means plus SEM. P values were calculated by two-tailed ratio paired Student's t test. ***, $P \leq 0.001$. n = 3 independent experiments using one NPC line. (c) Released infectious virus from siNT-treated or

siUPF1-treated, ZIKV-infected (MOI of 1) NPCs harvested at 48 hpi. Data are represented as means plus SEM. P values were calculated by unpaired Student's t test. **, $P \leq 0.01$. n = 3 independent experiments using one NPC line. (d) Representative confocal microscopy images of a ZIKV-infected, siNT-treated NPC or a ZIKV-infected, siUPF1-treated NPC with the nuclei stained with DAPI (blue) and ZIKV dsRNA foci stained with the anti-dsRNA mAb J2 (teal). 3D image rendering and reconstructed dsRNA foci were produced using the Imaris spot detection function. The scale bars represent 2 μm. (e) Number of dsRNA foci was averaged for each cell. Data are represented as means plus SEM. P values were calculated by two-tailed ratio paired Student's t test. ns, not significant. n = 3 independent experiments using two NPC lines, with 3 to 10 cells analyzed per condition for each experiment. (f) Measurements of dsRNA focus volume were averaged for each cell. Data are represented as means plus SEM. P values were calculated by two-tailed ratio paired Student's t test. ns, not significant. n = 3 independent experiments using two NPC lines, with 3 to 10 cells analyzed per condition for each experiment. (g) Infection rates of siNT-treated or siUPF1-treated, ZIKV-infected (MOI of 1) NPCs measured at 48 hpi. Fixed cells were subjected to the anti-DENV mAb 1.6D, which also recognizes the ZIKV envelope protein (56). Data are represented as means plus SEM. P values were calculated by two-tailed ratio paired Student's t test. **, $P \leq 0.01$. n = 3 independent experiments using two NPC lines. (h) Model of the interaction between the capsid protein of ZIKV and UPF1 of the NMD pathway.

FIG 1

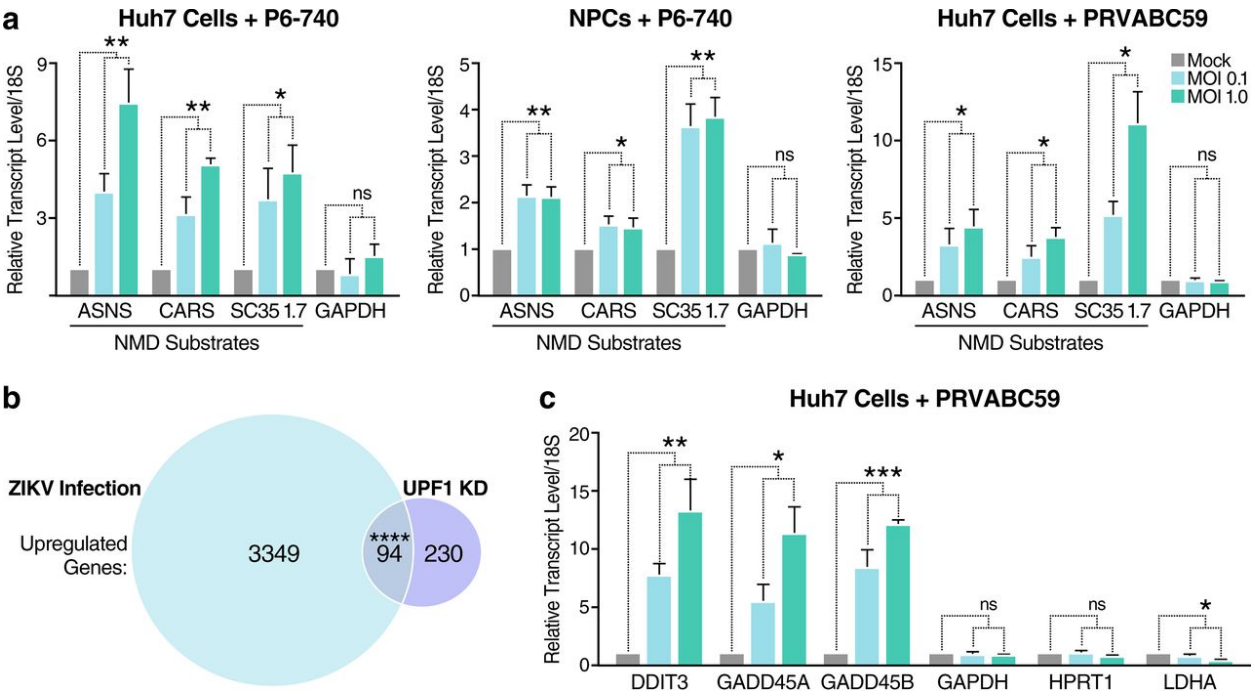

a
Huh7 Cells + P6-740
NPCs + P6-740
Huh7 Cells + PRVABC59
Relative Transcript Level/18S
Mock
MOI 0.1
MOI 1.0
ASNS
CARS
SC35 1.7
GAPDH
NMD Substrates
ns
b
ZIKV Infection
UPF1 KD
Upregulated Genes:
3349
****
94
230
c
Huh7 Cells + PRVABC59
Relative Transcript Level/18S
DDIT3
GADD45A
GADD45B
GAPDH
HPRT1
LDHA

FIG 2

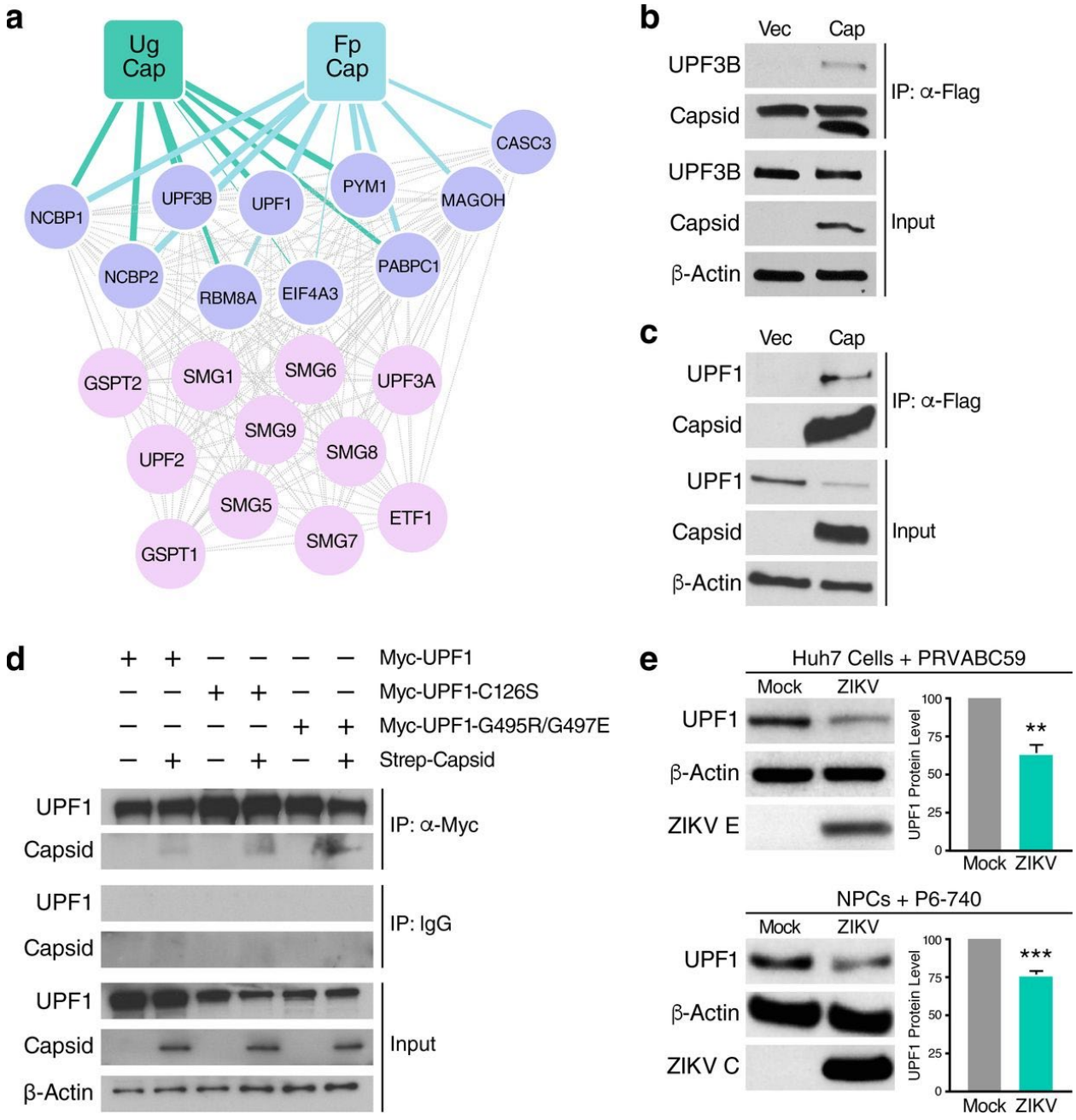

a
Ug Cap
Fp Cap
CASC3
NCBP1
UPF3B
UPF1
PYM1
MAGOH
NCBP2
RBM8A
EIF4A3
PABPC1
GSPT2
SMG1
SMG6
UPF3A
SMG9
SMG8
UPF2
SMG5
ETF1
SMG7
GSPT1
b
Vec
Cap
UPF3B
Capsid
IP: α-Flag
UPF3B
Capsid
Input
β-Actin
c
Vec
Cap
UPF1
Capsid
IP: α-Flag
UPF1
Capsid
Input
β-Actin
d
Myc-UPF1
Myc-UPF1-C126S
Myc-UPF1-G495R/G497E
Strep-Capsid
UPF1
Capsid
IP: α-Myc
UPF1
Capsid
IP: IgG
UPF1
Capsid
Input
β-Actin
e
Huh7 Cells + PRVABC59
Mock
ZIKV
UPF1
β-Actin
ZIKV E
UPF1 Protein Level
100
75
50
25
0
**
Mock
ZIKV
NPCs + P6-740
Mock
ZIKV
UPF1
β-Actin
ZIKV C
UPF1 Protein Level
100
75
50
25
0
***
Mock
ZIKV

FIG 3

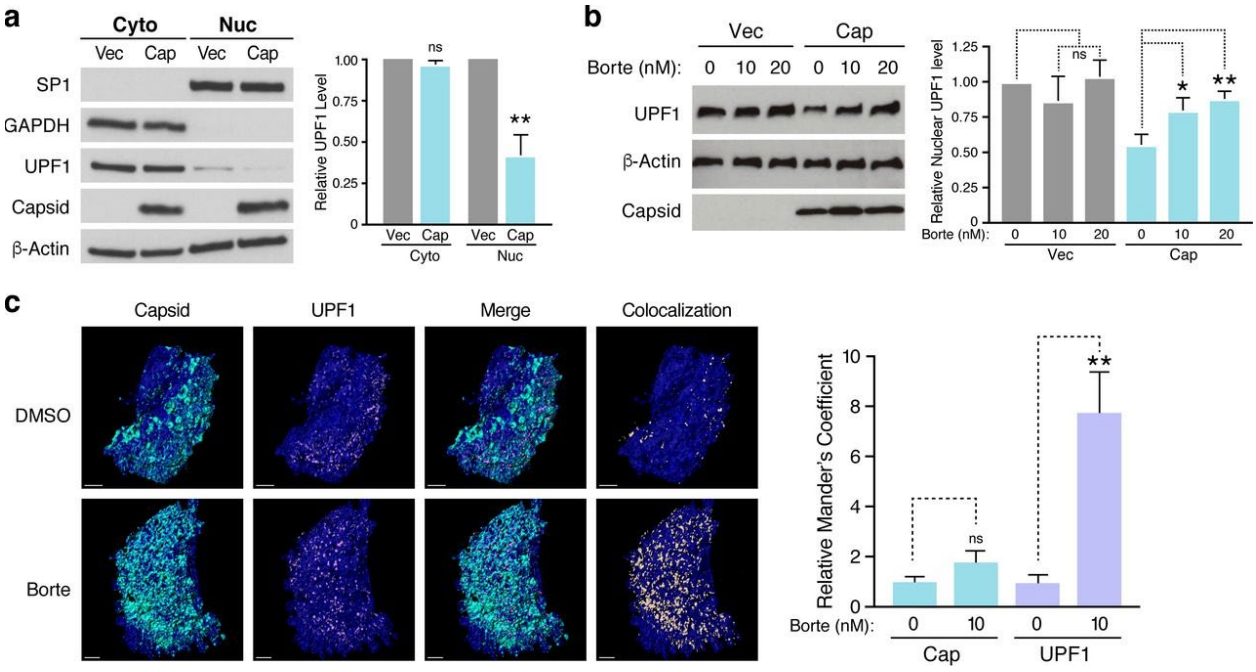
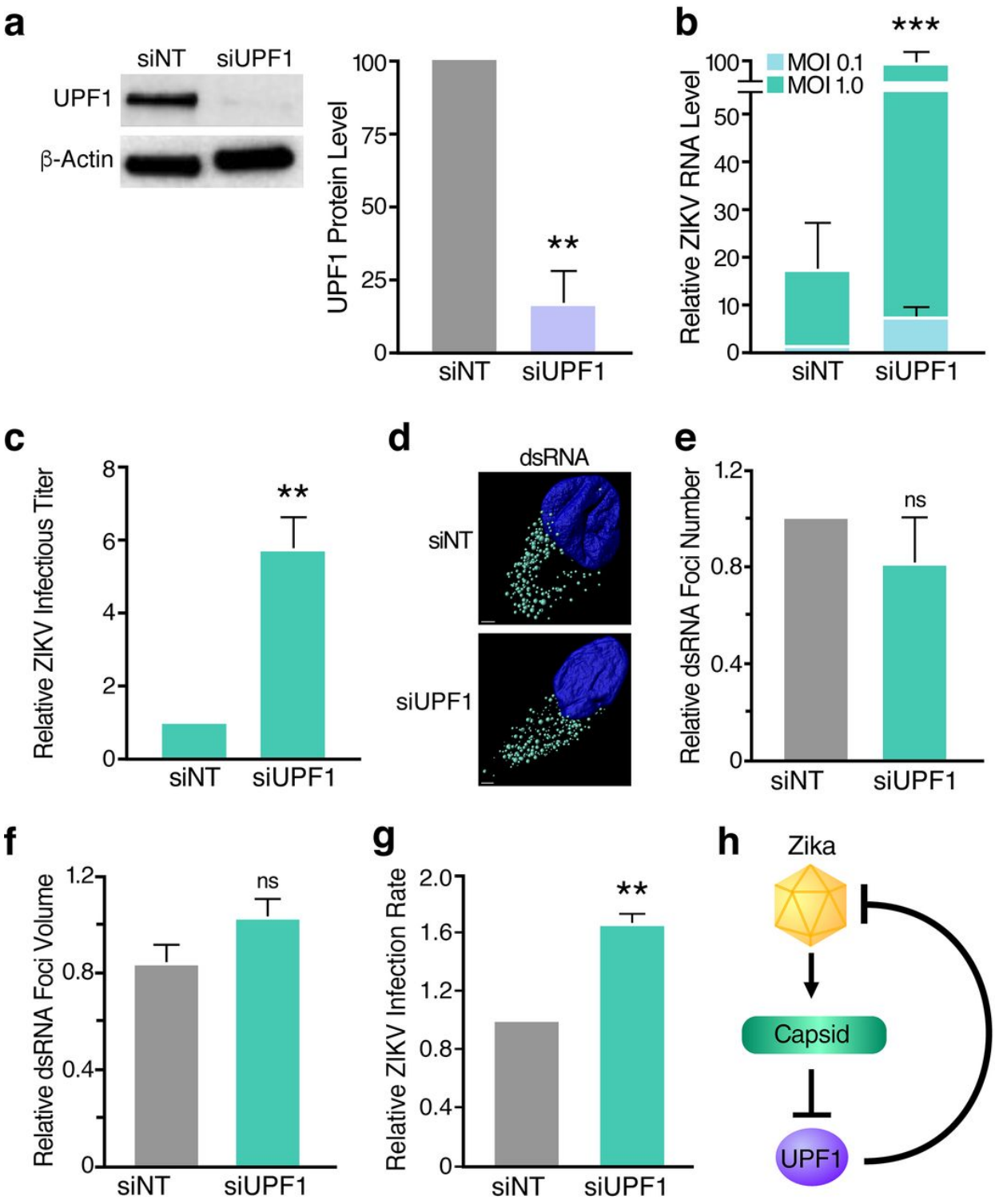

a
Cyto
Nuc
Vec
Cap
Vec
Cap
SP1
GAPDH
UPF1
Capsid
β-Actin
Relative UPF1 Level
1.00
0.75
0.50
0.25
0
ns
**
Vec
Cap
Vec
Cap
Cyto
Nuc
b
Vec
Cap
Borte (nM):
0
10
20
0
10
20
UPF1
β-Actin
Capsid
Relative Nuclear UPF1 level
1.25
1.00
0.75
0.50
0.25
0
ns
*
**
Borte (nM):
0
10
20
0
10
20
Vec
Cap
c
Capsid
UPF1
Merge
Colocalization
DMSO
Borte
Relative Mander's Coefficient
10
8
6
4
2
0
ns
**
Borte (nM):
0
10
0
10
Cap
UPF1

FIG 4


a
siNT
siUPF1
UPF1
β-Actin
UPF1 Protein Level
100
75
50
25
0
**
siNT
siUPF1
b
***
MOI 0.1
MOI 1.0
Relative ZIKV RNA Level
100
50
40
30
20
10
0
siNT
siUPF1
c
Relative ZIKV Infectious Titer
8
6
4
2
0
**
siNT
siUPF1
d
dsRNA
siNT
siUPF1
e
Relative dsRNA Foci Number
1.2
0.8
0.4
0
ns
siNT
siUPF1
f
Relative dsRNA Foci Volume
1.2
0.8
0.4
0
ns
siNT
siUPF1
g
Relative ZIKV Infection Rate
2.0
1.6
1.2
0.8
0.4
0
**
siNT
siUPF1
h
Zika
Capsid
UPF1

MATERIALS AND METHODS

Viruses and cells.

Two Asian lineage strains of ZIKV, P6-740 (ATCC VR-1845) and PRVABC59 (ATCC VR-1843), were used for all experiments. ZIKV stocks were propagated in Vero cells (ATCC), and titers were determined by plaque assays on Vero cells. Huh7 cells (ATCC), Huh7-Lunet cells (Ralf Bartenschlager, Heidelberg University), and Vero cells were maintained in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum (FBS), 2 mM l-glutamine, 100 U/ml penicillin, and 100 µg/ml streptomycin. HEK293T cells (ATCC) were maintained in DMEM/H21 medium supplemented with 10% FBS, 100 U/ml penicillin, 100 µg/ml streptomycin, and 1 mM sodium pyruvate or DMEM with 10% FBS, 2 mM l-glutamine, 100 U/ml penicillin, and 100 µg/ml streptomycin. Human iPSC-derived NPCs were generated and maintained as described previously (49). All of the human fibroblast cell lines used to generate iPSCs came from the Coriell Institute for Medical Research and Yale Stem Cell Center. The iPSCs used in these studies were the CTRL2493nXX, CS2518nXX, and Cs71iCTR-20nXX lines. CTRL2493nXX was derived from the parental fibroblast line ND31845 that was biopsied from a healthy female at 71 years of age. CS2518nXX was derived from the parental fibroblast line ND30625 that was biopsied from a healthy male at 76 years of age. CS71iCTR-20nXX was derived from the parental fibroblast line ND29971 that was biopsied from a female at 61 years of age. For virus infections, NPCs plated on Matrigel-coated (Corning) multiwell plates or Huh7 cells were infected with ZIKV at a multiplicity of infection (MOI) of 0.1 or 1 for 2 h at 37°C. Infected cells were harvested at 48 hpi for all analyses.

Affinity purification, mass spectrometry, and AP-MS scoring.

The ZIKV capsid open reading frames (ORFs) from the Ugandan 1947 strain MR 766 or the French Polynesian 2013 strain H/PF/2013 were cloned into pCDNA4_TO with a C-terminal 2xStrep II affinity tag for expression in human cells. The viral capsid proteins (three biological replicates), as well as GFP (two biological replicates) and empty vector (ten biological replicates) as negative controls, were expressed in HEK293T cells, and affinity purifications were performed as previously described (50). Briefly, clarified lysates were incubated with Strep-Tactin Superflow (IBA) overnight at 4°C. Proteins were eluted with 50 mM Tris (pH 7.5), 150 mM NaCl, and 1 mM EDTA containing 2.5 mM Desthiobiotin (IBA) for 30 min at 4°C. Lysates and affinity-purified eluates were analyzed by Western blotting and silver stain PAGE to confirm expression and purification. Purified protein eluates were digested with trypsin for LC-MS/MS analysis. Samples

were denatured and reduced in 2 M urea, 10 mM NH_4HCO_3, and 2 mM DTT for 30 min at 60°C and then alkylated with 2 mM iodoacetamide for 45 min at room temperature. Trypsin (Promega) was added at a 1:100 enzyme/substrate ratio and digested overnight at 37°C. Following digestion, samples were concentrated using C18 ZipTips (Millipore) according to the manufacturer's specifications. Peptides were resuspended in 15 µl of 4% formic acid and 3% ACN, and 1 to 2 µl of sample was loaded onto a 75-µm-ID column packed with 25 cm of Reprosil C18 1.9-µm, 120-Å particles (Dr. Maisch GmbH). Peptides were eluted into a Q-Exactive Plus (Thermo Fisher Scientific) mass spectrometer by gradient elution delivered by an Easy1200 nLC system (Thermo Fisher). The gradient was from 4.5% to 32% acetonitrile over 53 min. All MS spectra were collected with orbitrap detection, while the 20 most abundant ions were fragmented by higher energy collisional dissociation (HCD) and detected in the orbitrap. All data were searched against the Swiss-Prot Human protein sequences, combined with ZIKV sequences and GFP. Peptide and protein identification searches, as well as label-free quantitation, were performed using the MaxQuant data analysis algorithm, and all peptide and protein identifications were filtered to a 1% false-discovery rate (51, 52). SAINTq (53) was used to calculate the probability of bait-prey interactions for both Ugandan ZIKV capsid and French Polynesian ZIKV capsid against the negative controls, including GFP and empty vector, with protein intensities as input values. We applied a combined threshold of probability of interaction (AvgP) greater than 0.90 and a Bayesian false-discovery rate of less than 0.05.

Quantitative real-time reverse transcription-PCR (qRT-PCR).

Total cellular RNA was isolated from Huh7 cells and NPCs using the RNeasy Mini kit (Qiagen). cDNA was synthesized with oligo(dT)18 (Thermo Fisher Scientific) primers, random hexamer (Life Technologies) primers, and AMV reverse transcriptase (Promega). The cDNA was then used in SYBR green PCR master mix (Thermo Fisher Scientific) according to the manufacturer's instructions and analyzed by qPCR (Bio-Rad ABI 7900). The primers used for ASNS, CARS, SC35 1.7 (1.7 kb mRNA), GAPDH, HPRT1, LDHA, and 18S rRNA have been described previously (29). The additional primers used were as follows: ZIKV PRVABC59 forward primer, 5′-GAG ACG AGA TGC GGT ACA GG-3′; ZIKV PRVABC59 reverse primer, 5′-CGA CCG TCA GTT GAA CTC CA-3′; UPF1 forward primer, 5′-CTG CAA CGG ACG TGG AAA TAC-3′; UPF1 reverse primer, 5′-ACA GCC GCA GTT GTA GCA C-3′; DDIT3 forward primer, 5′- TG CTT CTC TGG CTT GGC TG-3′; DDIT3 reverse primer, 5′-GCT CTG GGA GGT GCT TGT GA-3′; GADD45A forward primer, 5′-GAG CTC CTG CTC TTG GAG AC-3′; GADD45A reverse primer, 5′-GCA GGA TCC TTC CAT TGA GA-3′; GADD45B forward primer, 5′-TGA CAA CGA CAT CAA

CAT C-3′; GADD45B reverse primer, 5′-GTG ACC AGA GAC AAT GCA G-3′. The relative levels of each transcript were normalized by the delta threshold cycle method to the abundance of 18S rRNA or GAPDH, with mock-infected cells or vector-transfected cells set at 1.

Western blot analysis.

Cells were lysed in RIPA lysis buffer (50 mM Tris-HCl [pH 8], 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, supplemented with Halt protease inhibitor cocktail [Thermo Fisher Scientific]) to obtain whole-cell lysates or lysed using the NE-PER nuclear and cytoplasmic extraction kit (Thermo Fisher Scientific) to obtain cytoplasmic and nuclear fractions. Proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes (Bio-Rad). Blots were incubated with the indicated primary antibody: anti-UPF3B (ab134566; Abcam), anti-UPF1 (12040; Cell Signaling Technology, Inc.), anti-ZIKV capsid (C) (GTX133304; GeneTex), anti-Flag (F7425; Sigma-Aldrich), anti-β-actin (A5316; Sigma-Aldrich), anti-ZIKV envelope (E) (GTX133314; GeneTex), anti-SP1 (sc-14027; Santa Cruz Biotechnology), anti-GAPDH (5174; Cell Signaling Technology, Inc.), anti-Myc tag (ab9106; Abcam), anti-Strep tag (ab18422, Abcam), and anti-p62 (ab56416, Abcam). Proteins were visualized by chemiluminescent detection with ECL and ECL Hyperfilm (Amersham). Differences in band intensity were quantified by densitometry using ImageJ.

Immunoprecipitations.

Cells were lysed in either RIPA lysis buffer or IP lysis buffer (150 mM NaCl, 50 mM Tris [pH 7.4], 1 mM EDTA, 0.5% NP-40 substitute, supplemented with Halt protease inhibitor cocktail [Thermo Fisher Scientific]) at 4°C and passed through a G23 needle. Clarified lysates were immunoprecipitated with Flag M2 agarose (Sigma), anti-Myc tag (ab9106; Abcam), or normal rabbit IgG (sc-2027; Santa Cruz Biotechnology) overnight, washed in lysis buffer, and resuspended in Laemmli buffer for SDS-PAGE. Western blot analysis of immunoprecipitated proteins was performed as described above.

Immunofluorescence.

Transfected Huh7-Lunet cells or infected NPCs were collected at 48 h and plated onto 22- by 22-mm no. 1.5 coverslips. Cells were fixed in 4% paraformaldehyde, permeabilized with 0.1% Triton X-100, and blocked in 3% bovine serum albumin. Cells were then immunostained with the indicated antibodies: anti-Strep Tag (ab184224; Abcam), anti-UPF1 (ab109363; Abcam), human anti-dengue virus (DENV) MAb 1.6D (Sharon Isern and Scott Michael, Florida Gulf Coast

University), which recognizes the ZIKV envelope protein, anti-dsRNA MAb J2 (SCICONS), and the appropriate fluorophore-conjugated secondary antibodies. Coverslips were mounted onto glass slides using Vectashield mounting medium with DAPI (Vector Laboratories) and analyzed by fluorescence microscopy (Zeiss Axio Observer ZI) or confocal microscopy (Zeiss LSM 880). For acquiring high-resolution images, cells were imaged on the Zeiss LSM 880 with Airyscan using a 20×/0.8 or 63×/1.4 M27 oil immersion objective. A total of 15 to 20 (20× objective) or 60 to 80 (63× objective) Z-slices were acquired every 0.88 µm or 0.15 µm, respectively. The resulting Z-stack was reconstructed and rendered in 3D using Imaris software (Bitplane). Viral dsRNA foci were reconstructed via the Imaris spot detection function, which provided an analysis of total number and mean volume of foci within a cell, for images acquired using the 20× objective. Strep-tagged ZIKV capsid, UPF1, and dsRNA channels acquired using the 63× objective were reconstructed using the Imaris surfaces package. The Imaris colocalization function was used to determine overlap of fluorescence. Thresholding for background fluorescence was determined by the Imaris automatic thresholding tool that utilizes the Costes approach (54). The thresholded Mander's correlation coefficient (MCC) measures the fraction of voxels with fluorescence positive for one channel that also contains fluorescence from another channel. The MCC is typically more appropriate for analysis of three-dimensional colocalization (55).

Statistical analysis.

Statistical differences between groups were analyzed using either a two-tailed unpaired Student's t test or a two-tailed ratio paired Student's t test as stated in the figure legends. Hypergeometrical tests were used to calculate the probability of an overlap in gene dysregulation between ZIKV-infected NPCs and UPF1-depleted cells and to calculate the probability of ZIKV capsid bait-prey interactions. Data are represented as means plus standard errors of the means (SEM). Statistical significance was defined as follows: *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$.

ACKNOWLEDGMENTS

We thank all members of the Ott laboratory, as well as Roman Camarda, Marius Walter, and Anna Maurer for helpful discussions and advice throughout the preparation of the manuscript. We thank Chia-Lin Tsou, the Gladstone Stem Cell Core, and Meredith Calvert from the Gladstone Microscopy Core for technical assistance, and Ralf Bartenschlager (Heidelberg University), Lynne Maquat (University of Rochester), and Sharon Isern and Scott Michael (Florida Gulf Coast University) for reagents. We are grateful to Veronica Fonseca and Lauren Weiser for

administrative support, John Carroll and Teresa Roberts for graphical design, and to Kathryn Claiborn, Eric Martens, and Gary Howard for editorial assistance.

This work was supported by NIH/NIAID F32AI112262 to P.S.S., NIH/NINDS R01 NS101996-01 to S.F., NIH/NIAID U19AI1186101 to N.J.K., DOD/DARPA HR0011-11-C-0094 (PROPHECY) to N.J.K., NIH/NIAID R01 AI097552 to M.O., BioFulcrum, and the James B. Pendleton Charitable Trust.

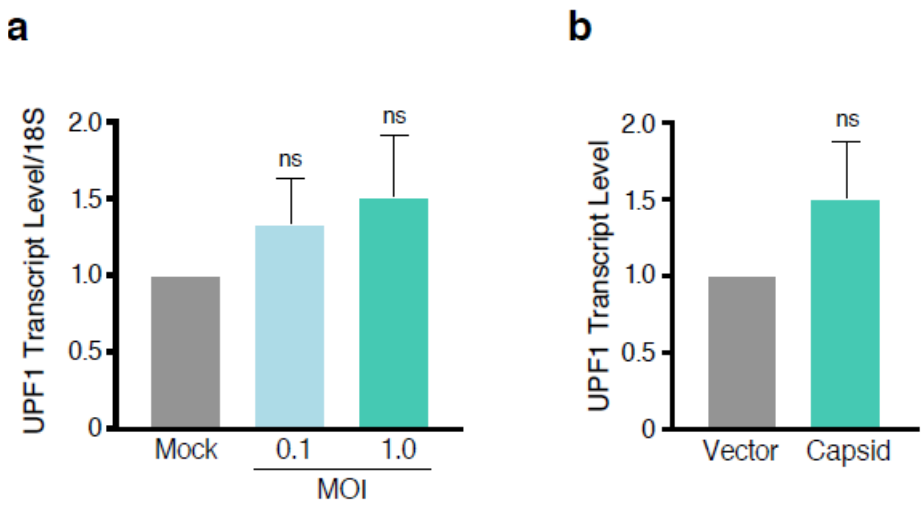


FIG S1

UPF1 is not transcriptionally downregulated during ZIKV infection or following ZIKV capsid overexpression. (a) UPF1 transcript levels from Huh7 cells mock infected or infected with ZIKV strain PRVABC59 at an MOI of 0.1 or 1 and harvested at 48 hpi. Data are represented as means plus SEM. P values were calculated by unpaired Student's t test. ns, not significant. n = 3 independent experiments. (b) UPF1 transcript levels from HEK293T cells transfected with vector or Strep-tagged ZIKV capsid (H/PF/2013, Asian lineage) and harvested at 48 hpt. Data are represented as means plus SEM. P values were calculated by unpaired Student's t test. ns, not significant. n = 3 independent experiments

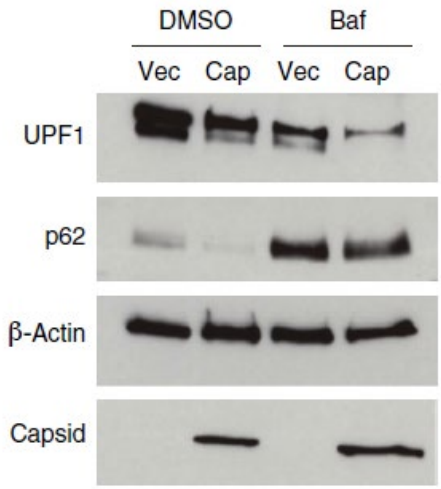


FIG S2

ZIKV capsid-induced UPF1 degradation is not dependent on autophagy. Western blot analysis of nuclear UPF1 levels in fractionated HEK293T cells transfected with vector or Flag-tagged ZIKV capsid for 48 h. Cells were treated with DMSO or the autophagy inhibitor bafilomycin A1 (Baf) (10 nM) for 24 h before harvest. Levels of p62, which is degraded by autophagy, were monitored to confirm autophagic inhibition following bafilomycin A1 treatment.

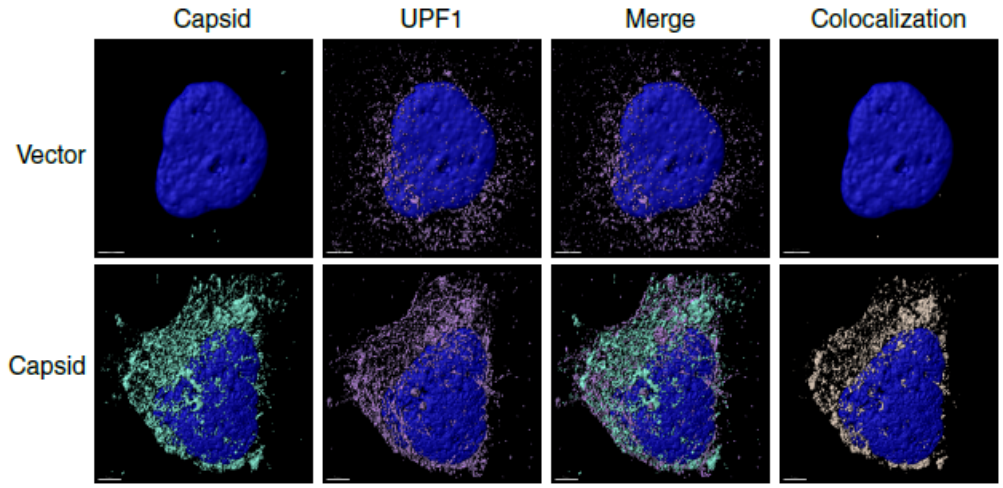


FIG S3

ZIKV capsid colocalizes with endogenous UPF1. Representative 3D confocal microscopy images of Huh7-Lunet cells transfected with vector or Strep-tagged ZIKV capsid. Cells were processed for immunostaining at 48 hpt and probed with antibodies against Strep tag (turquoise) and endogenous UPF1 (purple). DAPI (blue) was used to stain the nuclei. Each channel was reconstructed digitally for visualization of the 3D colocalization. The thresholded Mander's correlation coefficient for ZIKV capsid was 0.57 (n = 17), indicating that approximately 57% of the voxels positive for capsid fluorescence were also positive for UPF1 fluorescence. Scale bar represents 5 μm.

REFERENCES

1. Fauci AS, Morens DM. 2016. Zika virus in the Americas–yet another arbovirus threat. N Engl J Med 374:601–604. 10.1056/NEJMp1600297.

2. Cugola FR, Fernandes IR, Russo FB, Freitas BC, Dias JLM, Guimarães KP, Benazzato C, Almeida N, Pignatari GC, Romero S, Polonio CM, Cunha I, Freitas CL, Brandão WN, Rossato C, Andrade DG, Faria DdP, Garcez AT, Buchpigel CA, Braconi CT, Mendes E, Sall AA, Zanotto PM, Peron JPS, Muotri AR, Beltrão-Braga PCB. 2016. The Brazilian Zika virus strain causes birth defects in experimental models. Nature 534:267–271. 10.1038/nature18296.

3. Shao Q, Herrlinger S, Yang S, Lai F, Moore JM, Brindley MA, Chen J. 2016. Zika virus infection disrupts neurovascular development and results in postnatal microcephaly with brain damage. Development 143:4127–4136. 10.1242/dev.143768.

4. Souza BSF, Sampaio GLA, Pereira CS, Campos GS, Sardi SI, Freitas LAR, Figueira CP, Paredes BD, Nonaka CKV, Azevedo CM, Rocha VPC, Bandeira AC, Mendez-Otero R, Dos Santos RR, Soares MBP. 2016. Zika virus infection induces mitosis abnormalities and apoptotic cell death of human neural progenitor cells. Sci Rep 6:39775. 10.1038/srep39775.

5. Li C, Xu D, Xu Z, Ye Q, Hong S, Jiang Y, Liu X, Zhang N, Shi L, Qin C. 2016. Zika virus disrupts neural progenitor development and leads to microcephaly in mice. Cell Stem Cell 19:120–126. 10.1016/j.stem.2016.04.017.

6. Tang H, Hammack C, Ogden S, Wen Z, Qian X, Li Y, Yao B, Shin J, Zhang F, Lee E, Christian K, Didier R, Jin P, Song H, Ming G. 2016. Zika virus infects human cortical neural progenitors and attenuates their growth. Cell Stem Cell 18:587–590. 10.1016/j.stem.2016.02.016.

7. Rasmussen SA, Jamieson DJ, Honein MA, Petersen LR. 2016. Zika virus and birth defects–reviewing the evidence for causality. N Engl J Med 374:1981–1987. 10.1056/NEJMsr1604338.

8. Harris E, Holden KL, Edgil D, Polacek C, Clyde K. 2006. Molecular biology of flaviviruses. Novartis Found Symp 277:253.

9. Lindenbach BD, Rice CM. 2003. Molecular biology of flaviviruses. Adv Virus Res 59:23–61. 10.1016/S0065-3527(03)59002-9.

10. Oliveira ERA, Mohana-Borges R, de Alencastro RB, Horta BAC. 2017. The flavivirus capsid protein: structure, function and perspectives towards drug design. Virus Res 227:115–123. 10.1016/j.virusres.2016.10.005.

11. den Boon JA, Diaz A, Ahlquist P. 2010. Cytoplasmic viral replication complexes. Cell Host Microbe 8:77–85. 10.1016/j.chom.2010.06.010.

12. Colpitts TM, Barthel S, Wang P, Fikrig E. 2011. Dengue virus capsid protein binds core histones and inhibits nucleosome formation in human liver cells. PLoS One 6:e24365. 10.1371/journal.pone.0024365.

13. Slomnicki LP, Chung DH, Parker A, Hermann T, Boyd NL, Hetman M. 2017. Ribosomal stress and Tp53-mediated neuronal apoptosis in response to capsid protein of the Zika virus. Sci Rep 7:1–15.

14. Tsuda Y, Mori Y, Abe T, Yamashita T, Okamoto T, Ichimura T, Moriishi K, Matsuura Y. 2006. Nucleolar protein B23 interacts with Japanese encephalitis virus core protein and participates in viral replication. Microbiol Immunol 50:225–234. 10.1111/j.1348-0421.2006.tb03789.x.

15. Xu Z, Hobman TC. 2012. The helicase activity of DDX56 is required for its role in assembly of infectious West Nile virus particles. Virology 433:226–235. 10.1016/j.virol.2012.08.011.

16. Rawlinson SM, Moseley GW. 2015. The nucleolar interface of RNA viruses. Cell Microbiol 17:1108–1120. 10.1111/cmi.12465.

17. Peccarelli M, Kebaara BW. 2014. Regulation of natural mRNAs by the nonsense-mediated mRNA decay pathway. Eukaryot Cell 13:1126–1135. 10.1128/EC.00090-14.

18. Hug N, Longman D, Cáceres JF. 2016. Mechanism and regulation of the nonsense-mediated decay pathway. Nucleic Acids Res 44:1483–1495. 10.1093/nar/gkw010.

19. Han X, Wei Y, Wang H, Wang F, Ju Z, Li T. 2018. Nonsense-mediated mRNA decay: a 'nonsense' pathway makes sense in stem cell biology. Nucleic Acids Res 46:1038–1051. 10.1093/nar/gkx1272.

20. Silver DL, Watkins-Chow DE, Schreck KC, Pierfelice TJ, Larson DM, Burnetti AJ, Liaw H, Myung K, Walsh CA, Gaiano N, Pavan WJ. 2010. The exon junction complex component Magoh controls brain size by regulating neural stem cell division. Nat Neurosci 13:551–558. 10.1038/nn.2527.

21. Mao H, Pilaz L, McMahon JJ, Golzio C, Wu D, Shi L, Katsanis N, Silver DL. 2015. Rbm8a haploinsufficiency disrupts embryonic cortical development resulting in microcephaly. J Neurosci 35:7003–7018. 10.1523/JNEUROSCI.0018-15.2015.

22. Mao H, McMahon JJ, Tsai Y, Wang Z, Silver DL. 2016. Haploinsufficiency for core exon junction complex components disrupts embryonic neurogenesis and causes p53-mediated microcephaly. PLoS Genet 12:e1006282. 10.1371/journal.pgen.1006282.

23. Kurosaki T, Maquat LE. 2016. Nonsense-mediated mRNA decay in humans at a glance. J Cell Sci 129:461–467. 10.1242/jcs.181008.

24. Serquiña AKP, Das SR, Popova E, Ojelabi OA, Roy CK, Göttlinger HG. 2013. UPF1 is crucial for the infectivity of human immunodeficiency virus type 1 progeny virions. J Virol 87:8853–8861. 10.1128/JVI.00925-13.

25. Molleston JM, Cherry S. 2017. Attacked from all sides: RNA decay in antiviral defense. Viruses 9:2. 10.3390/v9010002.

26. Balistreri G, Bognanni C, Mühlemann O. 2017. Virus escape and manipulation of cellular nonsense-mediated mRNA decay. Viruses 9:24. 10.3390/v9010024.

27. Rigby RE, Rehwinkel J. 2015. RNA degradation in antiviral immunity and autoimmunity. Trends Immunol 36:179–188. 10.1016/j.it.2015.02.001.

28. Balistreri G, Horvath P, Schweingruber C, Zünd D, McInerney G, Merits A, Mühlemann O, Azzalin C, Helenius A. 2014. The host nonsense-mediated mRNA decay pathway restricts mammalian RNA virus replication. Cell Host Microbe 16:403–411. 10.1016/j.chom.2014.08.007.

29. Ramage HR, Kumar GR, Verschueren E, Johnson JR, Von Dollen J, Johnson T, Newton B, Shah P, Horner J, Krogan NJ, Ott M. 2015. A combined proteomics/genomics approach links hepatitis C virus infection with nonsense-mediated mRNA decay. Mol Cell 57:329–340. 10.1016/j.molcel.2014.12.028.

30. Tani H, Imamachi N, Salam KA, Mizutani R, Ijiri K, Irie T, Yada T, Suzuki Y, Akimitsu N. 2012. Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. RNA Biol 9:1370–1379. 10.4161/rna.22360.

31. Weischenfeldt J, Damgaard I, Bryder D, Theilgaard-Mönch K, Thoren LA, Nielsen FC, Jacobsen SEW, Nerlov C, Porse BT. 2008. NMD is essential for hematopoietic stem and

progenitor cells and for eliminating by-products of programmed DNA rearrangements. Genes Dev 22:1381–1396. 10.1101/gad.468808.

32. Li T, Shi Y, Wang P, Guachalla LM, Sun B, Joerss T, Chen Y, Groth M, Krueger A, Platzer M, Yang Y, Rudolph KL, Wang Z. 2015. Smg6/Est1 licenses embryonic stem cell differentiation via nonsense-mediated mRNA decay. EMBO J 34:1630–1647. 10.15252/embj.201489947.

33. Brazão TF, Demmers J, van IJcken W, Strouboulis J, Fornerod M, Romão L, Grosveld FG. 2012. A new function of ROD1 in nonsense-mediated mRNA decay. FEBS Lett 586:1101–1110. 10.1016/j.febslet.2012.03.015.

34. Chan W, Huang L, Gudikote JP, Chang Y, Imam JS, MacLean JA, Wilkinson MF. 2007. An alternative branch of the nonsense-mediated decay pathway. EMBO J 26:1820–1830. 10.1038/sj.emboj.7601628.

35. Nelson JO, Moore KA, Chapin A, Hollien J, Metzstein MM. 2016. Degradation of Gadd45 mRNA by nonsense-mediated decay is essential for viability. Elife 5:e12876. 10.7554/eLife.12876.

36. Jauhiainen A, Thomsen C, Strömbom L, Grundevik P, Andersson C, Danielsson A, Andersson MK, Nerman O, Rörkvist L, Ståhlberg A, Åman P. 2012. Distinct cytoplasmic and nuclear functions of the stress induced protein DDIT3/CHOP/GADD153. PLoS One 7:e33208. 10.1371/journal.pone.0033208.

37. Salvador JM, Brown-Clay JD, Fornace AJ. 2013. Gadd45 in stress signaling, cell cycle control, and apoptosis. Adv Exp Med Biol 793:1–19. 10.1007/978-1-4614-8289-5_1.

38. Gehring NH, Lamprinaki S, Kulozik AE, Hentze MW. 2009. Disassembly of exon junction complexes by PYM. Cell 137:536–548. 10.1016/j.cell.2009.02.042.

39. Chamieh H, Ballut L, Bonneau F, Le Hir H. 2008. NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. Nat Struct Mol Biol 15:85–93. 10.1038/nsmb1330.

40. Varsally W, Brogna S. 2012. UPF1 involvement in nuclear functions. Biochem Soc Trans 40:778–783. 10.1042/BST20120052.

41. Klema VJ, Padmanabhan R, Choi KH. 2015. Flaviviral replication complex: coordination between RNA synthesis and 5′-RNA capping. Viruses 7:4640–4656. 10.3390/v7082837.

42. Maisel M, Herr A, Milosevic J, Hermann A, Habisch H, Schwarz S, Kirsch M, Antoniadis G, Brenner R, Hallmeyer-Elgner S, Lerche H, Schwarz J, Storch A. 2007. Transcription profiling of adult and fetal human neuroprogenitors identifies divergent paths to maintain the neuroprogenitor cell state. Stem Cells 25:1231–1240. 10.1634/stemcells.2006-0617.

43. Li H, Saucedo-Cuevas L, Regla-Nava J, Chai G, Sheets N, Tang W, Terskikh A, Shresta S, Gleeson J. 2016. Zika virus infects neural progenitors in the adult mouse brain and alters proliferation. Cell Stem Cell 19:593–598. 10.1016/j.stem.2016.08.005.

44. Nickless A, Bailis JM, You Z. 2017. Control of gene expression through the nonsense-mediated RNA decay pathway. Cell Biosci 7:1–12. 10.1186/s13578-017-0153-7.

45. Cortese M, Goellner S, Acosta EG, Neufeldt CJ, Oleksiuk O, Lampe M, Haselmann U, Funaya C, Schieber N, Ronchi P, Schorb M, Pruunsild P, Schwab Y, Chatel-Chaix L, Ruggieri A, Bartenschlager R. 2017. Ultrastructural characterization of Zika virus replication factories. Cell Rep 18:2113–2123. 10.1016/j.celrep.2017.02.014.

46. Grant A, Ponia SS, Tripathi S, Balasubramaniam V, Miorin L, Sourisseau M, Schwarz MC, Sánchez-Seco MP, Evans MJ, Best SM, García-Sastre A. 2016. Zika virus targets human STAT2 to inhibit type I interferon signaling. Cell Host Microbe 19:882–890. 10.1016/j.chom.2016.05.009.

47. von Mikecz A. 2006. The nuclear ubiquitin-proteasome system. J Cell Sci 119:1977–1984. 10.1242/jcs.03008.

48. Freire JM, Veiga AS, Conceição TM, Kowalczyk W, Mohana-Borges R, Andreu D, Santos NC, Da Poian AT, Castanho MARB. 2013. Intracellular nucleic acid delivery by the supercharged dengue virus capsid protein. PLoS One 8:e81450. 10.1371/journal.pone.0081450.

49. HD iPSC Consortium. 2017. Developmental alterations in Huntington's disease neural cells and pharmacological rescue in cells and mice. Nat Neurosci 20:648–660. 10.1038/nn.4532.

50. Jäger S, Gulbahce N, Cimermancic P, Kane J, He N, Chou S, D'Orso I, Fernandes J, Jang G, Frankel AD, Alber T, Zhou Q, Krogan NJ. 2011. Purification and characterization of HIV-human protein complexes. Methods 53:13–19. 10.1016/j.ymeth.2010.08.007.

51. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. 2014. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics 13:2513–2526. 10.1074/mcp.M113.031591.

52. Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26:1367–1372. 10.1038/nbt.1511.

53. Teo G, Koh H, Fermin D, Lambert J, Knight JDR, Gingras A, Choi H. 2016. SAINTq: scoring protein-protein interactions in affinity purification – mass spectrometry experiments with fragment or peptide intensity data. Proteomics 16:2238–2245. 10.1002/pmic.201500499.

54. Costes SV, Daelemans D, Cho EH, Dobbin Z, Pavlakis G, Lockett S. 2004. Automatic and quantitative measurement of protein-protein colocalization in live cells. Biophys J 86:3993–4003. 10.1529/biophysj.103.038422.

55. Dunn KW, Kamocka MM, McDonald JH. 2011. A practical guide to evaluating colocalization in biological microscopy. Am J Physiol Cell Physiol 300:C723–C742. 10.1152/ajpcell.00462.2010.

56. Paul LM, Carlin ER, Jenkins MM, Tan AL, Barcellona CM, Nicholson CO, Michael SF, Isern S. 2016. Dengue virus antibodies enhance Zika virus infection. Clin Transl Immunol 5:e117. 10.1038/cti.2016.72.

Chapter 5

Nuclear Accumulation of Host Transcripts during Zika Virus Infection and UPF1 knockdown

Kristoffer E. Leon, Ryan A. Flynn,# Mir M. Khalid,# Krystal A. Fontaine, Thong T Nguyen, G. Renuka Kumar, Camille R. Simoneau, Sakshi Tomar, David Jimenez-Morales, Mariah Dunlap, Julia Kaye, Priya S. Shah, Steven Finkbeiner, Nevan J. Krogan, Carolyn Bertozzi, Jan E. Carette, Melanie Ott.*

Submitted.

Biorxiv, 2020

DOI: 10.1101/2020.12.03.410837

#These authors contributed equally *Lead contact

Nuclear Accumulation of Host Transcripts during Zika Virus Infection and UPF1 knockdown

Kristoffer E. Leon[1,2,3,4], Ryan A. Flynn[5#], Mir M. Khalid[1#], Krystal A. Fontaine[1], Thong T Nguyen[1], G. Renuka Kumar[1], Camille R. Simoneau[1,2,4], Sakshi Tomar[1,6], David Jimenez-Morales[1,7], Mariah Dunlap[1], Julia Kaye[1], Priya S. Shah[8], Steven Finkbeiner[1,9,10], Nevan J. Krogan[1,11,12], Carolyn Bertozzi[5,13], Jan E. Carette[14], Melanie Ott[1,2*]

[1]J. David Gladstone Institutes, San Francisco, CA
[2]Department of Medicine, University of California, San Francisco, CA
[3]Medical Scientist Training Program, University of California, San Francisco, CA
[4]Biomedical Sciences Graduate Program, University of California, San Francisco, CA
[5]Department of Chemistry and Stanford ChEM-H, Stanford University, Stanford, CA
[6]National Cancer Institute, National Institutes of Health, Bethesda, MD
[7]Division of Cardiovascular Medicine, Department of Medicine, Stanford University, Stanford, CA
[8]Departments of Chemical Engineering and Microbiology and Molecular Genetics, University of California, Davis, CA
[9]Center for Systems and Therapeutics and Taube/Koret Center for Neurodegenerative Disease Research, San Francisco, CA
[10]Departments of Neurology and Physiology, University of California, San Francisco, CA
[11]Quantitative Biosciences Institute (QBI), University of California, San Francisco, CA
[12]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA
[13]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA
[14]Department of Microbiology and Immunology, Stanford University, Stanford, CA

#RAF and MMK contributed equally.

[*]Corresponding Author and Lead Contact
Address correspondence to: melanie.ott@gladstone.ucsf.edu
Gladstone Institutes
1650 Owens Street
San Francisco, CA 9415

Abstract

Zika virus (ZIKV) infects fetal neural progenitor cells (NPCs) causing severe neurodevelopmental disorders including microcephaly. We previously showed that levels of up-frameshift protein 1 (UPF1), an RNA-binding protein regulating transcript degradation, are decreased in infected NPCs, specifically in the cell nucleus. Here we find that in infected NPCs approximately a third of UPF1 target transcripts no longer bind UPF1, consistent with its diminished expression. Notably, these transcripts are not altered in abundance but in subcellular partitioning with mRNAs accumulating in the nucleus in infected or UPF1 knockdown cells. This causes decreased protein expression of FREM2, a protein required for proper NPC function. Our results point to a central function of UPF1 in mRNA transport and implicate the perturbation of this function in neurodevelopmental defects associated with ZIKV infection.

Introduction

Zika virus (ZIKV) is a mosquito-borne, enveloped virus with a positive sense, single stranded RNA genome (Musso and Gubler, 2016). ZIKV is a member of the *Flaviviridae* family that includes Dengue virus (DENV), West Nile virus (WNV) and Hepatitis C virus (HCV). In 2015, an outbreak of ZIKV in Brazil was linked to a dramatic increase in the number of infants born with microcephaly (Fauci and Morens, 2016; Pierson and Diamond, 2018). It was subsequently shown that *in utero*, ZIKV infects neural progenitor cells (NPCs), resulting in neurodevelopmental delays that ultimately cause a range of birth defects including microcephaly, ocular damage, and contractures, collectively known as congenital Zika syndrome (Costa et al., 2016; Cugola et al., 2016; Shao et al., 2016; Souza et al., 2016; Tang et al., 2016). NPCs are critical for brain development as they differentiate into the glial and neuronal cells that compose the majority of the brain parenchyma (Martínez-Cerdeño and Noctor, 2018). Proposed molecular mechanisms by which ZIKV disrupts NPC function and differentiation focus on multiple cellular pathways including centrosomal organization, autophagy, apoptosis and unfolded protein response pathways (Link et al., 2019; Ojha et al., 2018; Saade et al., 2020; Shah et al., 2018; Wen et al., 2019). However, it remains unknown how ZIKV manipulates multiple cellular pathways at once to cause such widespread developmental reprogramming.

We and others previously showed that ZIKV infection suppresses the host nonsense-mediated mRNA decay (NMD) pathway (Fontaine et al., 2018; Li et al., 2019). NMD is an RNA quality control mechanism that targets faulty host transcripts for degradation and acts as an antiviral pathway on many viral species, particularly single-stranded RNA viruses (Balistreri et al., 2017; Leon and Ott, 2020). Different models of the NMD pathway have been described, but the exon-junction complex (EJC)-mediated NMD pathway is the best defined (Kurosaki and Maquat, 2016; Kurosaki et al., 2019). This pathway is classically triggered by premature termination codons in mRNAs, but can be also be initiated by normal 3'UTRs (Gehring et al., 2003).

EJCs are normally loaded onto mRNAs in the nucleus (Popp and Maquat, 2013); the RNAs are then exported to the cytoplasm to undergo translation, leading to proper displacement of the EJCs (Gehring et al., 2009). Non-displaced EJCs recruit the master regulator and helicase up-frameshift protein 1 (UPF1), which induces a recruitment cascade activating NMD-mediated degradation of the faulty transcript (Loh et al., 2013; Lykke-Andersen et al., 2014). However, the NMD pathway and UPF1 also target non-faulty mRNAs, especially those with long and GC-rich 3' untranslated regions (UTRs), a feature found in many viral RNAs (Imamachi et al., 2017; Kebaara and Atkin, 2009; Peccarelli and Kebaara, 2014; Toma et al., 2015). UPF1 has been implicated in the regulation of a wide range of biological processes including other RNA decay

pathways besides NMD (Kim and Maquat, 2019), telomere maintenance (Chawla et al., 2011), protein ubiquitination (Feng et al., 2017) and RNA export (Ajamian et al., 2015; Singh et al., 2019).

Because knockdown of UPF1 enhances replication of many RNA viruses including ZIKV, WNV, DENV, Rous Sarcoma virus, Potato virus X, Pea Enation Mosaic virus 2, and Turnip Crinkle virus, NMD is recognized as a *bone fide* antiviral restriction pathway, often targeted by viruses including ZIKV for inactivation (Leon and Ott, 2020). ZIKV inactivation of the antiviral activity of NMD involves the ZIKV capsid protein directly targeting UPF1 (Fontaine et al., 2018; Li et al., 2019). Similar interactions of capsid proteins with the NMD pathway also occur with other flaviviruses (Li et al., 2019). However, the finding that ZIKV capsid expression selectively downregulates UPF1 expression in the host nucleus was unexpected as viral RNA replication occurs exclusively in the host cytoplasm (Fontaine et al., 2018). It remained unclear how downregulation of nuclear UPF1 would support viral replication.

Here we show that nuclear mRNA export is a major UPF1 function that is disrupted during ZIKV infection. ZIKV infection or UPF1 knockdown resulted in polyadenylated transcript accumulation in the nucleus, causing decreased protein levels of FREM2, a UPF1 target, and consequent perturbation of NPC differentiation. We propose that by targeting nuclear UPF1 and trapping host mRNAs in the nucleus, ZIKV has evolved a mechanism to “shut off” host mRNA function while promoting translation of its own proteins. This mechanism describes a new central role for UPF1 in mRNA export connected to many cellular pathways associated with neurodevelopment.

Results

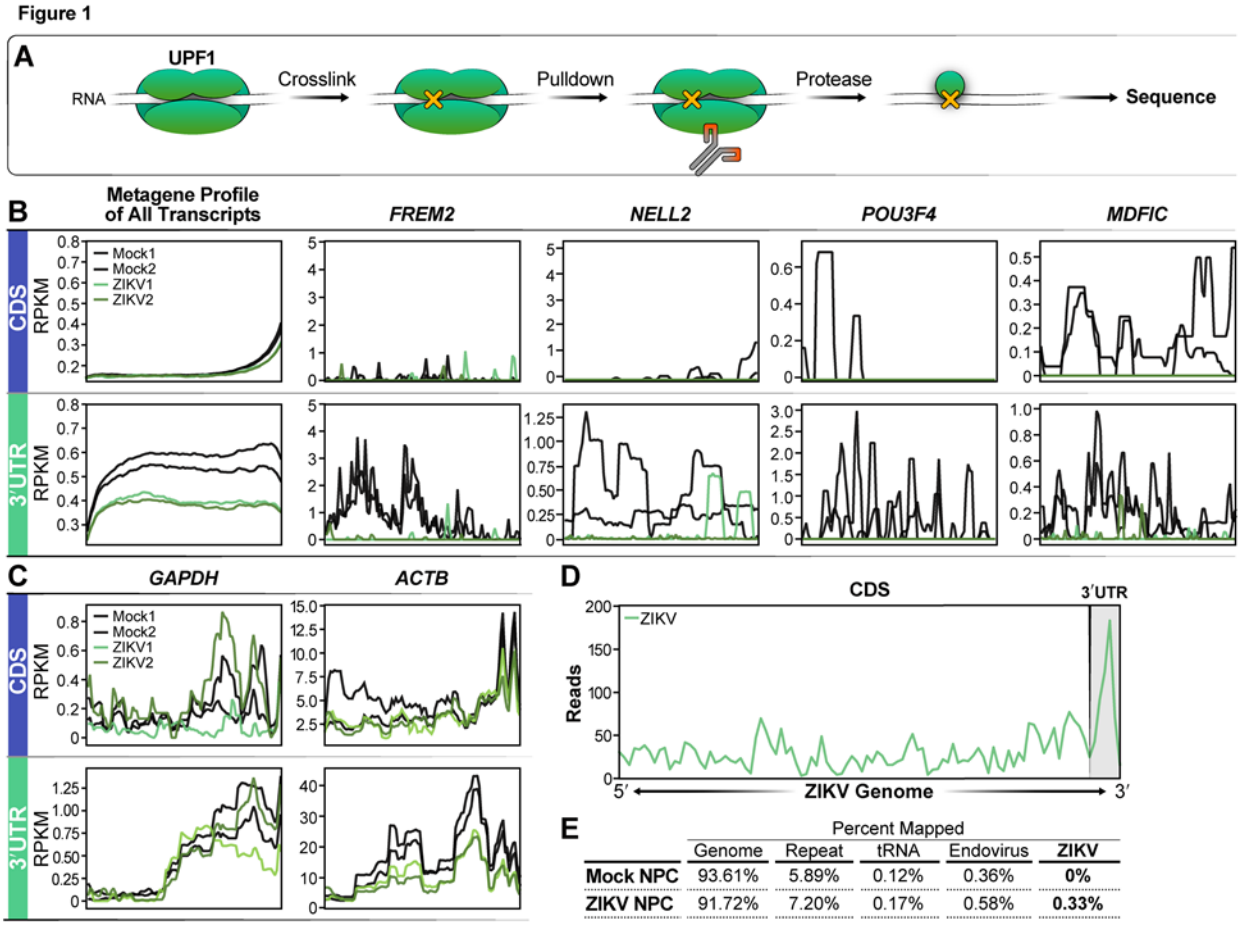
ZIKV infection decreases UPF1 interaction with the 3’UTR of host transcripts

While ZIKV inhibits NMD by capsid-mediated degradation of UPF1, the functional impact of UPF1 loss is still unknown. We first aimed to comprehensively define the UPF1-RNA interactions in human induced pluripotent stem cell-derived NPCs by performing infrared crosslinking immunoprecipitation and RNA sequencing (irCLIP-Seq) (Zarnegar et al., 2016). UPF1 is an RNA helicase and ATPase found associated with many transcripts (Fiorini et al., 2015). To identify differences in UPF1 occupancy during infection, NPCs were infected with ZIKV (isolate PRVABC59) at an MOI of 1 for 48 hours followed by UV-crosslinking of transcripts with proteins and immunoprecipitation of UPF1 (**Figure 1A**). Mock-infected cultures served as controls. RNA-protein complexes were separated by SDS-PAGE, and mass spectrometry was performed on the excised bands to confirm UPF1 enrichment (**Figure S1A, Supplemental File 1**) before RNA was extracted and submitted to next-generation sequencing (**Figure 1A**, **Supplemental File 2**).

In mock-infected NPCs, UPF1 bound 6778 transcripts, predominantly at the 3'UTR, as previously reported (Kebaara and Atkin, 2009; Toma et al., 2015). In ZIKV-infected NPCs, this number decreased to 4557 transcripts, with a marked decrease in occupancy at the 3'UTR (**Figure 1B**). Steady-state transcripts such as glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and Beta-Actin (ACTB) maintained UPF1 interaction (**Figure 1C**).

In infected samples, UPF1 also interacted with ZIKV RNA, mostly within the viral 3'UTR, supporting the notion of an antiviral function of UPF1 through RNA binding and degradation (**Figure 1D**). We speculated that viral RNA could act as a "sponge" sequestering UPF1 away from host RNAs and explaining decreased UPF1 occupancy of host transcripts during infection. However, quantification of UPF1-bound reads mapping to host vs. viral RNAs showed that viral reads only accounted for 0.33% of all mapped reads while the majority of sequences isolated with UPF1 pulldown mapped to the host transcriptome. This excluded a "sponge" effect of viral RNA for UPF1 **(Figure 1E).**

Figure 1



	Genome	Repeat	tRNA	Endovirus	ZIKV
Mock NPC	93.61%	5.89%	0.12%	0.36%	**0%**
ZIKV NPC	91.72%	7.20%	0.17%	0.58%	**0.33%**

Figure 1: UPF1-host transcript interactions are decreased during ZIKV infection of NPCs
A) irCLIP schematic describing the workflow to obtain sequencing data. After 48 hrs of ZIKV infection, UPF1 and RNA are crosslinked using UV light, followed by UPF1 pulldown and protease degradation to expose the UPF1-bound RNA for sequencing. **B)** Metagene profile of all transcripts sequenced from the irCLIP experiment. The graphs show Reads Per Kilobase of transcript, per Million mapped reads (RPKM) values for positions in the CDS and the 3'UTR. Experiment was produced from 2 biological replicates. Representative metagene plots (FREM2, NELL2, POU3F4, MDFIC) of loci found to have a loss of UPF1 interaction. **C)** A metagene plot of GAPDH, a transcript resistant to UPF1 degradation, and ACTB are shown as negative controls. **D)** Metagene plot of reads mapping to the ZIKV genome, with the 3'UTR marked. **E)** Tabular breakdown of read map percentages from the UPF1-CLIP experiment.

Selective reduction of UPF1 target transcripts in the cytoplasm of infected NPCs
As UPF1 is central to NMD, we examined whether transcripts that lost UPF1 occupancy would be stabilized. We performed whole transcriptome sequencing of ZIKV-infected NPCs followed by differential expression analysis (**Supplemental File 3**). Consistent with previous RNA-seq studies of ZIKV-infected NPCs (Li et al., 2016; Liu et al., 2019), genes associated with the interferon response to infection were predominantly upregulated (**Figure 2A, Supplemental Figure 2B**). In contrast, the abundance of transcripts which lost UPF1 occupancy upon infection was not significantly changed (**Figure 2B, Supplemental File 2**). This indicates that likely only a fraction of UPF1-occupied transcripts is subject to degradation, underscoring the functional relevance of UPF1 functions outside of NMD.

Metascape analysis of transcripts with stable abundance after loss of UPF1 occupancy showed enrichment for neural and neurodevelopmental functions, such as axon guidance, highlighting the importance of UPF1 for neurodevelopment as previously indicated (Jaffrey and Wilkinson, 2018; Stoeckli, 2018) **(Figure 2C**). NMD and UPF1 have been implicated in regulation of axon guidance, including specific localization patterns in different neuron types (Colak et al., 2013).

Next, we examined whether transcript localization was changed when UPF1 occupancy was lost. This was prompted by previous reports that indicated that UPF1 translocates between the nucleus and cytoplasm (Varsally and Brogna, 2012), and our previous observation that UPF1 is selectively downregulated in the nucleus upon infection (Fontaine et al., 2018). ZIKV- and Mock-infected NPCs were fractionated into cytoplasmic and nuclear compartments and RNA

sequencing as well as differential expression analysis was performed for each fraction (**Supplemental File 4,5)**.

We confirmed successful fractionation by comparing the $\log_2$ fold change of known cytoplasmic (*GAPDH*) and nuclear (*ANRIL*) transcripts. *GAPDH* was 3.2-fold enriched in the cytoplasmic fraction and ANRL 2.3-fold in the nuclear fraction, as expected (**Supplemental Figure 2A**). Upon ZIKV infection, we identified 585 and 312 significantly upregulated mRNAs in the cytoplasm and nucleus, respectively (**Figure 2C**). These upregulated transcripts were predominantly interferon response genes (**Supplemental Figure 2D**). In contrast, downregulated genes were overwhelmingly cytoplasmic, with 912 genes downregulated in the cytoplasm compared to 22 in the nuclear fraction (**Figure 2E**). Moreover, of the top 53 genes with the most significant decrease in UPF1 interaction upon infection, all but 6 were exclusively downregulated in the cytoplasm (**Figure 2F**). These results show a selective loss of UPF1 target mRNAs in the cytoplasm of infected cells. As total abundance of these transcripts was unchanged, we considered the possibility that they accumulate in the nucleus upon ZIKV infection.

Figure 2

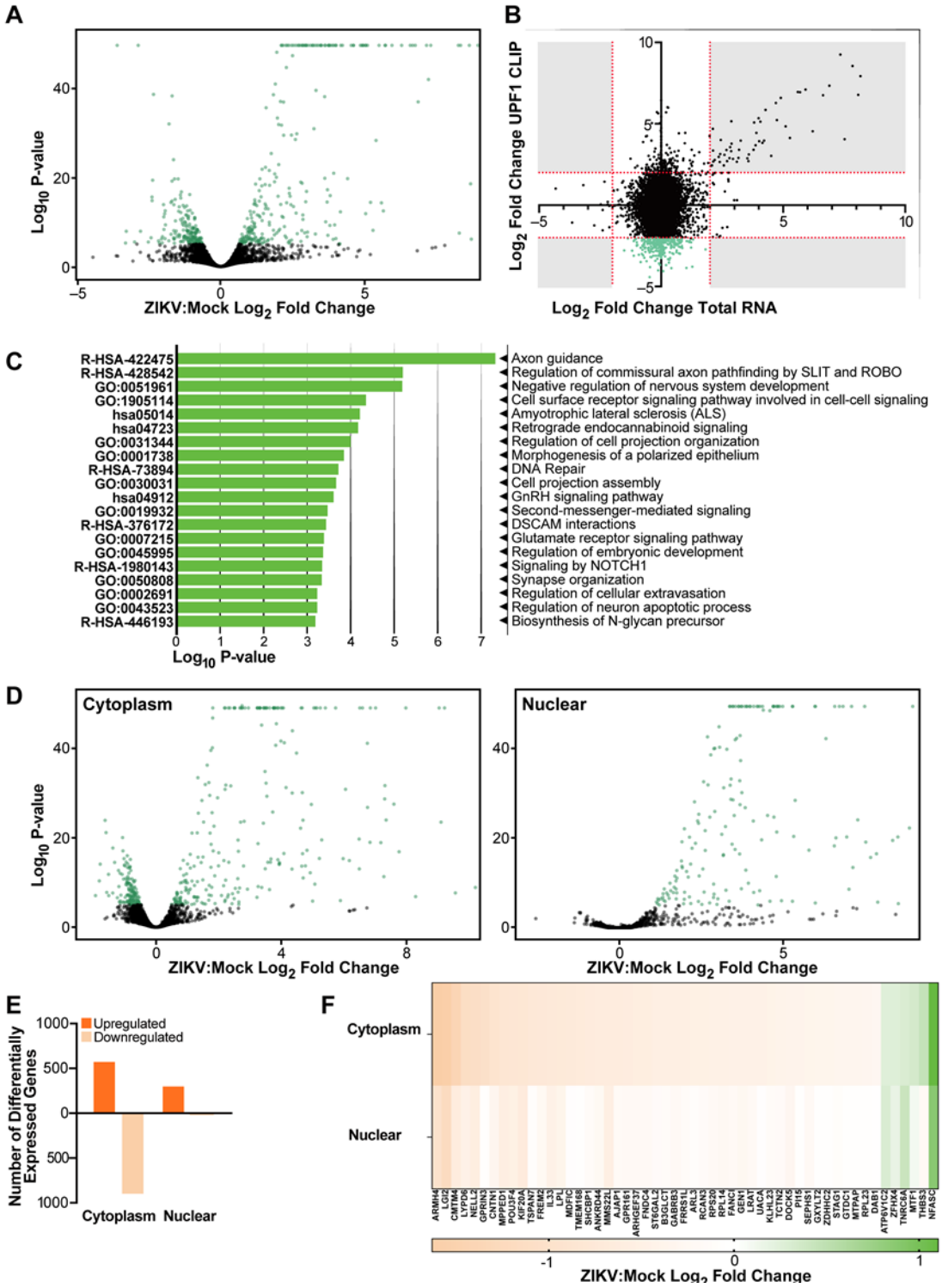

A
Log10 P-value
ZIKV:Mock Log2 Fold Change
B
Log2 Fold Change UPF1 CLIP
Log2 Fold Change Total RNA
C
R-HSA-422475
R-HSA-428542
GO:0051961
GO:1905114
hsa05014
hsa04723
GO:0031344
GO:0001738
R-HSA-73894
GO:0030031
hsa04912
GO:0019932
R-HSA-376172
GO:0007215
GO:0045995
R-HSA-1980143
GO:0050808
GO:0002691
GO:0043523
R-HSA-446193
Axon guidance
Regulation of commissural axon pathfinding by SLIT and ROBO
Negative regulation of nervous system development
Cell surface receptor signaling pathway involved in cell-cell signaling
Amyotrophic lateral sclerosis (ALS)
Retrograde endocannabinoid signaling
Regulation of cell projection organization
Morphogenesis of a polarized epithelium
DNA Repair
Cell projection assembly
GnRH signaling pathway
Second-messenger-mediated signaling
DSCAM interactions
Glutamate receptor signaling pathway
Regulation of embryonic development
Signaling by NOTCH1
Synapse organization
Regulation of cellular extravasation
Regulation of neuron apoptotic process
Biosynthesis of N-glycan precursor
Log10 P-value
D
Cytoplasm
Nuclear
Log10 P-value
ZIKV:Mock Log2 Fold Change
E
Upregulated
Downregulated
Number of Differentially Expressed Genes
Cytoplasm
Nuclear
F
Cytoplasm
Nuclear
ZIKV:Mock Log2 Fold Change

Figure 2: **Loss of UPF1 interaction during ZIKV infection causes mRNA downregulation in the cytoplasm**

A) Volcano plot of whole transcriptome RNA sequencing of ZIKV infected NPCs compared to Mock, n=2 biological replicates. **B)** X-axis shows ZIKV: Mock log_2 fold change in whole transcriptome sequencing, Y-Axis shows ZIKV:Mock log_2 fold change from the UPF1 CLIP experiment. The green transcripts are >2 log_2 fold change in the UPF1-CLIP and <2 log2 fold change in total RNA. **C)** Metascape analysis of transcripts identified to have a significant loss of UPF1 interaction in ZIKV infected cells compared to Mock. **D)** Volcano plot of sequencing from ZIKV-infected NPCs compared to Mock fractionated into cytoplasmic and nuclear fractions, n=2 biological replicates. **E)** Number of differentially expressed genes from sequencing of ZIKV-infected NPCs compared to Mock fractionated into cytoplasmic and nuclear fractions **F)** A heatmap showing log_2 fold changes of transcripts identified in the CLIP experiment in cytoplasmic and nuclear fractions.

ZIKV infection, ZIKV capsid expression or UPF1 knockdown results in accumulation of host transcripts in the nucleus

To test the hypothesis that transcripts were selectively accumulating in the nucleus in ZIKV-infected cells, we analyzed global transcript localization using RNAscope analysis with a polyA tail probe. We first optimized the technology in Huh7-Lunet cells, a hepatoma cell line frequently used to study flavivirus infection (Vicenti et al., 2018), and NPCs, both infected with ZIKV at an MOI of 1. In both cell types, polyadenylated RNA fluorescence was significantly increased in the nucleus upon infection, supporting the model that mRNA location rather than abundance is perturbed in ZIKV-infected cells (**Figure 3A,B**). To understand if this effect was UPF1-mediated, we leveraged the previously demonstrated ability of ZIKV capsid to specifically degrade nuclear UPF1 (Fontaine et al., 2018), and used Huh7-Lunet cells expressing ZIKV capsid protein in a tetracycline-inducible manner. Induction of capsid expression by addition of doxycycline led to reduced levels of nuclear UPF1 (**Figure 3C**). This also led to a similar accumulation of polyadenylated RNAs in the cell nucleus, mirroring the effect observed in ZIKV-infected cells (62% nuclear fluorescence in vector vs 78% in capsid overexpressing cells) (**Figure 3D**). Furthermore, siRNA-mediated knockdown of UPF1 in NPCs also resulted in retention of polyadenylated RNAs in the nucleus (36% nuclear fluorescence in control vs 46% in siUPF1 NPCs) (**Figure 3E**, **Supplemental Figure 3**). Collectively, these data indicate that UPF1 regulates mRNA transport, a function perturbed upon nuclear UPF1 degradation during ZIKV infection.

Figure 3

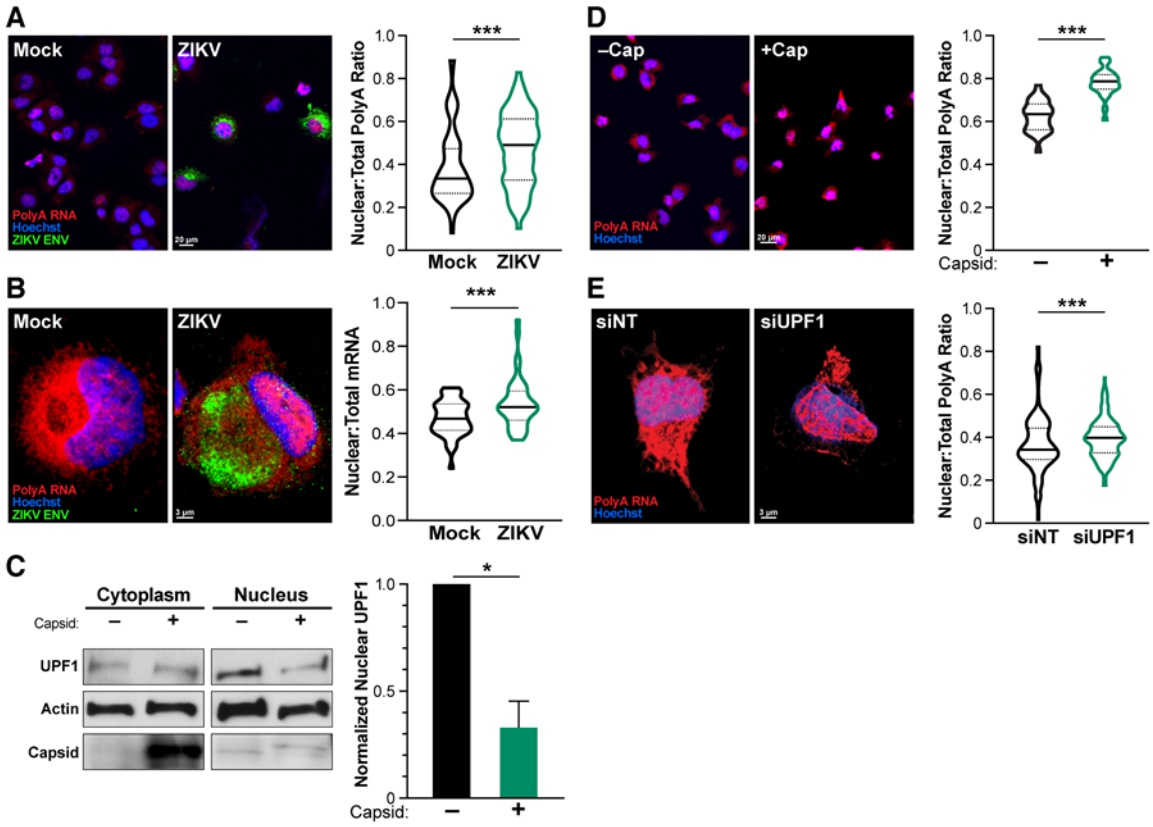


Figure 3: ZIKV-mediated degradation of UPF1 leads to mRNA retention in the nucleus.
A and B) During ZIKV infection (ZIKV ENV, green), polyA RNA (red) is shifted toward the nucleus (Hoechst, blue) in Huh7-Lunet cells (A) and NPCs (B). Statistics produced by a Linear Mixed Model. 3 biological replicates, n=25 cells (A) or 15 cells (B) per condition per replicate. **C)** Tetracycline-inducible Capsid expression in Huh7-Lunet cells was used to degrade nuclear UPF1. Leptomycin B (LMB) was used to increase UPF1 degradation. Statistics performed by Student's t-test, n=3, and representative western blot shown. **D)** Capsid overexpression in Huh7-Lunet cells results in an increased ratio of polyA RNAs (red) in the nucleus. N=3 biological replicates, 25 cells per condition per replicate. **E)** Knockdown of UPF1 in NPCs results in an increased ratio of polyA

RNAs (red) in the nucleus. NPCs were treated with siNT and siUPF1 for 96 hours. Statistics produced by a Linear Mixed Model. n=3 biological replicates, 15 cells per condition per replicate. *, P ≤ 0.05; **, P ≤ 0.01; ***, P ≤ 0.001. Error bars are SEM.

Depletion of UPF1 leads to nuclear retention of *FREM2* mRNA, depletion of FREM2 protein, and alterations to NPC differentiation markers

Next, we focused on *FREM2*, the transcript with the largest fold decrease in UPF1 interaction in the irCLIP studies (**Figure 1B**). FREM2 is an extracellular matrix protein involved in cell-cell interactions that is important in many developmental pathways including tissue and vascular morphogenesis (Pavlakis et al., 2011; Timmer et al., 2005). The *FREM2* transcript was specifically downregulated in the cytoplasm of infected NPCs (**Figure 2G**). We confirmed selective cytoplasmic downregulation of *FREM2* mRNA expression upon infection by qPCR (**Figure 4A**). Using RNAscope, we found enrichment of the *FREM2* transcript in the nucleus of cells treated with UPF1-targeting siRNAs as compared to cells treated with non-targeting control siRNAs, where it was found in both the cytoplasm and the nucleus (**Figure 4B**). FREM2 protein levels were significantly decreased in UPF1 siRNA-treated cells (42% decrease in siUPF1 compared to the siNT control), supporting the model that nuclear mRNA retention leads to reduced translation of UPF1 target transcripts during ZIKV infection (**Figure 4C**).

To determine the functional consequences of reduced FREM2 expression, we used siRNAs to knockdown FREM2 in NPCs (**Supplemental Figure 4).** FREM2 knockdown increased the percentage of cells expressing the neuronal marker βIII-Tubulin by 44%, and decreased the percentage of cells expressing the pluripotency marker Sox2 by 51% (**Figure 4D**). These results indicate that by preventing FREM2 nuclear export and protein expression, ZIKV infection may promote premature neuronal differentiation of NPCs, thus reducing the number of NPCs that determine brain and ultimately head size.

Figure 4

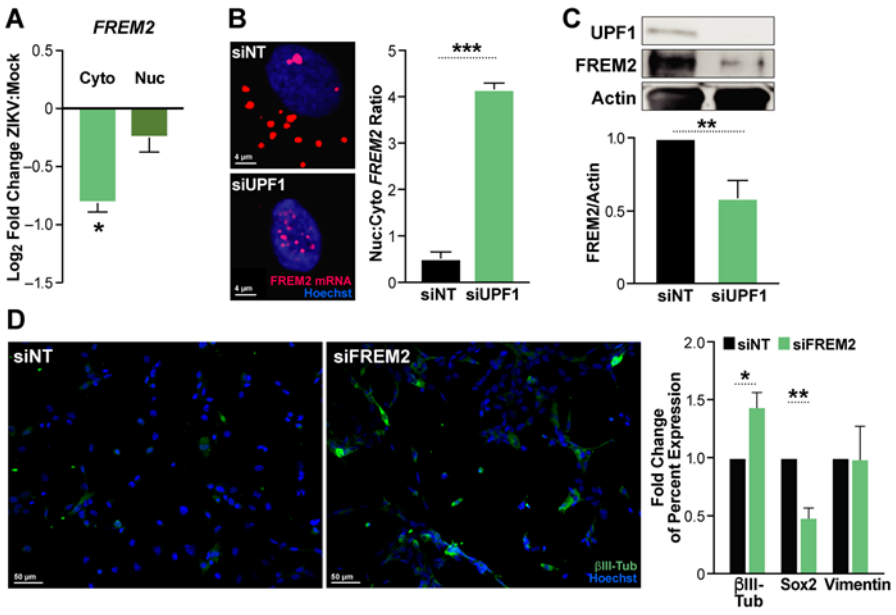


Figure 4: UPF1 knockdown leads to retention of *FREM2* mRNA and decreased protein production, which alters differentiation markers.

A) qPCR of *FREM2* in cytoplasmic and nuclear fractions in ZIKV infected NPCs. **B)** RNAscope of *FREM2* in UPF1 knockdown NPCs. NPCs were treated with siRNAs for 96 hours prior to harvesting for microscopy. Number of nuclear puncta compared to cytoplasmic as calculated by Imaris. 3 biological replicates, 10 cells per biological replicate per condition averaged. Statistics produced by Student's t-test. **C)** Western blot of siNT and siUPF1 treated NPCs. NPCs were treated with siRNAs for 96 hours prior to harvesting for microscopy. Densitometric analyses of FREM2 were performed using ImageJ to quantify relative band intensities. **D)** siNT and siFREM2 treated NPCs stained for βIII-Tubulin, Sox2 and Vimentin. NPCs were treated with siRNAs for 7 days prior to analysis. Statistics produced by Student's t-test. *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$. Error bars are SEM.

Discussion

In this study, we describe a potential host “shut-off” mechanism by which ZIKV infection prevents host mRNA export from the nucleus, a phenotype described in other viral infections, but not seen previously with flaviviruses (Fortes et al., 1994; Kumar and Glaunsinger, 2010; Kuss et al., 2013; Sakuma et al., 2014). Many transcripts lost UPF1 occupancy during ZIKV infection, especially in the 3’UTR, and this loss of UPF1 was associated with altered transcript localization and only a minor effect on transcript abundance. This underscores the significance of UPF1 functions outside its well-described function in the NMD pathway. The accumulation of mRNAs in the nucleus induced by ZIKV infection could be recapitulated both by ZIKV capsid expression, which is known to degrade nuclear UPF1, and by UPF1 knockdown, implicating UPF1’s nuclear localization in this function. Lastly, we show that when UPF1 is unable to export *FREM2* mRNA from the nucleus, FREM2 protein abundance decreases, supporting the model that viral targeting of nuclear UPF1 decreases translation of cellular mRNAs which may disrupt neural differentiation. The majority of these experiments were performed in human NPCs, the natural target cells of ZIKV in fetuses, providing a physiological context for the pathways and transcripts disrupted by ZIKV infection and potentially leading to congenital Zika syndrome.

Our studies in NPCs and Huh7 hepatoma cells indicate that at least in these cell types, UPF1 plays a prominent role in nuclear export. How this function is regulated and whether it involves known UPF1 interacting proteins such as the EJC or known RNA export pathways such as CRM-1, remains to be determined. Two other studies have linked UPF1 to nuclear export. The first study found that HIV-1 required UPF1 to export viral RNA from the nucleus of HeLA cells (Ajamian et al., 2015). The second study found that UPF1 translocates between the nucleus and cytoplasm in *Drosophila*, and UPF1 depletion disrupts mRNA export from the nucleus (Singh et al., 2019). The finding that ZIKV has evolved a unique function for its capsid protein to interfere with UPF1’s role in human mRNA export underscores the central role of UPF1 in this process, at least in the two cell types that we studied here.

While the precise mechanisms of UPF1-mediated nuclear export are not yet clear, two findings underscore its significance: 1) most transcripts were not upregulated in NPCs after loss of UPF1 interaction, which points to another major function of UPF1 in cells other than NMD, and 2) many polyadenylated transcripts were retained in the nucleus upon UPF1 knockdown, demonstrating that the effect of UPF1 on mRNA export is widespread. Additional studies are needed to determine whether specific features of these mRNAs, such as the 3’UTR sequence or structure, underlie their targeting for export by UPF1 (Carmody and Wente, 2009).

Our studies confirm that UPF1 binds ZIKV RNA consistent with its restrictive role through direct degradation. We map this binding to the ZIKV 3'UTR, similar to what is observed with host mRNAs. Interestingly, a non-coding subgenomic flavivirus RNA (sfRNA) is derived from the 3' UTR of ZIKV and is known to be resistant to XRN1 degradation (Göertz et al., 2018). The production of sfRNA is shared among flaviviruses and antagonizes the interferon response (Schuessler et al., 2012). Our data are consistent with UPF1 potentially regulating sfRNA production or interacting with ZIKV sfRNAs, which could affect ZIKV replication independently from NMD or could play a role in NMD evasion.

We speculate that the widespread nuclear retention of host mRNAs caused by the disruption of UPF1 function contributes to the neurodevelopmental defects seen during fetal ZIKV infections. Since UPF1 depletion impacts many mRNAs, it may represent a central mechanism explaining pleotropic effects of ZIKV infection on diverse cellular pathways. We pursued *FREM2* as a prominent UPF1 target, and show that it is downregulated at the protein level when UPF1 is unable to perform its role as a nuclear mRNA export regulator. FREM2 is both a member of the FRAS/FREM complex and regulates its formation (Pavlakis et al., 2011). The FRAS/FREM complex is found in cellular basement membranes and is expressed differentially during development (Timmer et al., 2005). FREM2 itself is important for proper development of the eye and FREM2 mutation is associated with Fraser syndrome, in which cryptopthalamos (a congenital defect where the eyes are covered completely by skin and often associated with small or missing eyeballs) is commonly seen (Slavotinek and Tifft, 2002). Developmental defects in the eyes are also seen in children with congenital Zika syndrome, supporting the possibility that retention of the *FREM2* transcript in the nucleus could be directly involved in pathogenesis (de Paula Freitas et al., 2017; Ventura and Ventura, 2018).

Nuclear retention is a new host shutoff mechanism previously not described for flaviviruses. Flaviviruses have been reported to cause host shutoff by translational repression (Roth et al., 2017). We confirm that a single viral protein, the capsid protein, is sufficient to degrade nuclear UPF1 and cause nuclear mRNA retention. Notably, capsid is one of few ZIKV proteins that can localize to the cell nucleus (Oliveira et al., 2017), which is independent from its function as a structural component of the virion (Mohd Ropidi et al., 2020). Our study highlights the significance of nuclear UPF1, which has been associated with cell cycle progression, DNA replication, telomere maintenance, and mRNA release (Singh et al., 2019; Varsally and Brogna, 2012) but is overall less well studied than the cytoplasmic form. The finding that ZIKV has evolved a mechanism to selectively target nuclear UPF1 and disrupt its mRNA export function

underscores the central role that this form of UPF1 plays in both viral defense and neural development.

Materials and Methods

Cell Culture and Viruses

Human iPSC-derived NPCs were generated and maintained as described previously (Consortium, 2017). The human fibroblast cell line used to generate iPSCs came from the Coriell Institute for Medical Research and Yale Stem Cell Center. The iPSCs used in these studies was the CTRL2493n17 line. CTRL2493n17 was derived from the parental fibroblast line ND31845 that was biopsied from a healthy female at 71 years of age. iPSCs were cultured and maintained in complete mTESR (StemCell Techologies, Vancouver, CA). NPCs were differentiated and maintained using EFH media (Stemline Neural Stem Cell Media [Sigma Aldrich, St. Louis, MO], EGF [R&D Biosystems, Minneapolis, MN], rhFGF basic [R&D Biosystems]. Heparin Sulfate [Sigma Aldrich]). NPCs were dissociated and plated onto Matrigel-coated plates (Corning, Corning, NY) prior to infecting with ZIKV (MOI of 1) or treating with siRNAs. Experiments were harvested 48 hours post infection and up to 7 days after siRNA treatment. Huh7-Lunet cells (Ralf Bartenschlager, Heidelberg University), and Vero cells were maintained in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum (FBS), 2 mM l-glutamine, 100 U/ml penicillin, and 100 µg/ml streptomycin.

N-terminally Strep TagII tagged ZIKV Capsid was cloned to pLVX-TetOne-Puro Vector (Clontech, Mountainview, CA, Cat: 631849) using BamHI & EcoRI cut sites. We used a 2nd generation lentiviral system to generate a stably expressing ZIKV Capsid Huh7-Lunet cell line. After transduction, Huh7-Lunet cells were selected with puromycin (2ug/ml) for one weeks and later clonally expanded. Clones with the highest and specific expression were used for the experiment.

The strain PRVABC59 of ZIKV (ATCC, Manassas, VA, VR-1843) was used for all experiments. ZIKV stocks were propagated in Vero cells (ATCC), and titers were determined by plaque assays on Vero cells. ZIKV infections were performed by adding viral inoculum to DMEM with 2% FBS or EFH followed by a two-hour incubation at 37C with a rock every 15 minutes. After infection was completed, inoculum was aspirated and then fresh DMEM with 10% FBS or EFH was added to the cells. Infected cells were cultured for 48 hours prior to harvesting for all sequencing and IF experiments.

Antibodies and other reagents

Primary antibodies used were anti-UPF1 (Bethyl laboratories, Montgomery, TX, A300-38A, CST, Newburyport, MA, 12040S and Abcam, Cambridge, UK, ab109363), anti-FREM2 (Invitrogen,

Carlsbad, CA, PA5-20982), anti-FLAG (Abcam, ab18230), anti-Actin (CST, 4967S), anti-GFAP (Abcam, ab53554), anti-Nestin (Abcam, ab22035) and anti-Beta III tubulin (Abcam, ab18207). Secondary antibodies used include goat anti-rabbit Alexa 488 (Invitrogen, A-11008), goat anti-mouse Alexa 488 (Invitrogen, A-11001), goat anti-rabbit Alexa 594 (Invitrogen, A-11012), goat anti-mouse Alexa 594 (Invitrogen, A-11005), donkey anti-goat Alexa 647 (Invitrogen, A-21447), donkey anti-rabbit Alexa 488 (Invitrogen, A-21206), donkey anti-mouse Alexa 594 (Invitrogen, A-21203). The RNAScope Multiplex Fluorescent V2 Assay (ACD, Newark, CA, 323100) was used with RNAscope Probes include polyA RNA (ACD, 318631) and FREM2 (ACD, 482841). Opal570 (Akoya Biosciences, Marlborough, MA, FP1488001KT) was used for visualization of RNAscope probes. Accell siRNA was used for knockdown of FREM2 (Dharmacon, Lafayette, CO, E-021693-00-0010) and UPF1 (Dharmacon, E-011763-00-0010) according to manufacturer's instructions. Non-targeting siRNAs used were cat. D-001910-10-20 (Dharmacon). Huh7-Lunets were treated with 60 ng/mL of Leptomycin B (Cayman Chemical) for 16 hours.

Infrared Crosslinking and Immunoprecipitation

irCLIP was performed as in Zarnegar et al. 2016. HeLa cells grown as described above and UV crosslinked to a total of 0.35 J/cm^2. Whole-cell lysates were generated in CLIP lysis buffer (50 mM HEPES, 200 mM NaCl, 1 mM EDTA, 10% glycerol, 0.1% NP-40, 0.2% Triton X-100, 0.5% N-lauroylsarcosine) and briefly sonicated using a probe-tip Branson sonicator to solubilize chromatin. Each experiment was normalized for total protein amount, typically 1 mg, and partially digested with RNase A (ThermoFisher Scientific, Waltham, MA, EN0531) for 10 minutes at 37°C and quenched on ice. UPF1 (Bethyl laboratories, A300-38A) IP's were performed using 15 μg of each antibody with 50 μL Protein G Dynabeads (ThermoFisher Scientific), for 8 hours at 4°C on rotation. Samples were washed sequentially in 1 mL for 1 minute each at 25°C: 1× high stringency buffer (15 mM Tris-HCl, pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% Triton X-100, 1% sodium deoxycholate, 120 mM NaCl, 25 mM KCl), 1× high salt buffer (15 mM Tris-HCl pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% Triton X-100, 1% sodium deoxycholate, 1 M NaCl), 2× NT2 buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM $MgCl_2$, 0.05% NP-40). After the NT2 wash, RNA-protein complexes were dephosphorylated with T4 PNK (NEB) for 45 minutes in an Eppendorf Thermomixer at 37°C, 15 seconds 1400rpm, 90 seconds of rest in a 30 μL reaction, pH 6.5, containing 10 units of T4 PNK, 0.1 μL SUPERase-IN (ThermoFisher Scientific), and 6 μL of PEG-400 (16.7% final). Dephosphorylated RNA-protein complexes were then rinsed once with NT2 buffer and 3'-end ligated with T4 RNA Ligase 1 (NEB, Ipswich, MA) overnight in an Eppendorf Thermomixer at 16°C, 15 seconds 1400rpm, 90 seconds of rest in a 60 μL reaction containing 10

units T4 RNA Ligase, 1.5 pmol pre-adenylated-IR800-3'biotin DNA-adapter, 0.1 μL SUPERase-IN, and 6 μL of PEG400 (16.7% final). The following day, samples were again rinsed once with 500 μL NT2 buffer and resuspended in 30μL of 20 mM DTT, 1x LDS (ThermoFisher Scientific) in NT2 buffer. Samples were heated to 75°C for 10 min, and released RNA-protein complexes were separated on 4-12% Bis-Tris SDS-PAGE (1.0mm X 12 well) at 200V for 45 min. Resolved RNP complexes were wet-transferred to nitrocellulose at 550 mA for 45 minutes at 4°C.

Nitrocellulose membranes were imaged using an Odyssey CLx scanner (LiCor, Lincoln, Nebraska), RBP-RNA complexes were excised using scalpels, and RNA was recovered by adding 0.1 mL of Proteinase K reaction buffer (100 mM Tris, pH 7.5, 50 mM NaCl, 1 mM EDTA, 0.2% SDS) and 5 μL of 20mg/mL Proteinase K (ThermoFisher Scientific). Proteins were digested for 60 minutes at 50°C in an Eppendorf Thermomixer. Next, 200 μL of saturated-phenol-chloroform, pH, 6.7 was added to each tube and incubated for 10 minutes at 37°C in an Eppendorf Thermomixer, 1400 rpm. Tubes were briefly centrifuged and the entire contents transferred to a 2 mL Heavy Phase Lock Gel (5Prime, South San Francisco, CA, 2302830). Samples were centrifuged for 2 minutes at >13000 rpm. The aqueous layer was re-extracted with 1 mL of chloroform (inverting 10 times to mix; no vortexing) in the same 2 mL Phase Lock Gel tube and centrifuged for 2 minutes at >13000 rpm. The aqueous layer was then transferred to a new 2 mL Heavy Phase Lock Gel tube and extracted again with an additional 1 mL of chloroform. After 2 minutes centrifugation at >13000 rpm, the aqueous layer was transferred to a siliconized 1.5 mL tube and precipitated overnight at -20°C by addition of 10 μL 5M NaCl, 3 μL Linear Polyacrylamide (ThermoFisher Scientific) and 0.8 mL 100% ethanol. RNA fragments were pelleted at >13000 rpm for 45 minutes at 4°C, washed once with 1 mL of ice cold 75% ethanol and air dried.

RNA pellets were resuspended in 12 μL water 1 μL of 3 μM cDNA and 1 μL of 10mM dNTPs and heated to 70°C for 5 minutes then rapidly cooled to 4°C. cDNA Master Mix (4 μL 5x Super Script IV (SSIV) Buffer, 1 μL 100mM DTT, 1 μL SSIV, 6 μL total) was added to the annealed RNA and incubated for 30 minutes at 55°C. cDNA:RNA hybrids were captured by addition of 5 μL of MyOne Streptavidin C1 Dynabeads (ThermoFisher Scientific) that had been rinsed and suspended in 50 μL of Biotin-IP buffer (100mM Tris, pH 7.5, 1M NaCl, 1mM EDTA, 0.1% Tween), and end over end rotation for 45 minutes at room temperature. Beads were placed on a 96-well magnet and washed sequentially twice with 100 μL of Biotin IP buffer and 100 μL ice-cold 1xPBS. Beads were resuspended in 10 μL of cDNA elution buffer (8.25 μL water, 1 μL of 1 μM P3 short oligo, and

0.75 μl of 50 mM $MnCl_2$) and heated to 95°C for 10 minutes, ramp 0.1 degree/second to 60°C forever. Next 5 μL of circularization reaction buffer was added (3.3 μL water, 1.5 μL 10x Circligase-II buffer, and 0.5 μL of Circligase-II (Epicentre, Madison, WI)). cDNA was circularized for 2 hours at 60°C. cDNA was purified with 30 μL of AMPure XP beads (Beckman Coulter, Pasadena, CA) and 75 μL of isopropanol. Samples were incubated for 20 minutes at 25°C, washed twice with 100 μL 80% ethanol, air dried for 5 minutes, and eluted in 14 μL of water. Elution took place at 95°C for 3 minutes and the eluent was immediately transferred to a 96-well magnet. Eluted cDNA was transferred to a new PCR tube containing 15 μL of 2X Phusion HF-PCR Master Mix (NEB), 0.5 μL of 30 μM P3/P6 PCR1 oligo mix and 0.5 μl of 15x SYBR Green I (ThermoFisher Scientific). Real-time quantitative PCR was performed: 98°C 2 min, 15 cycles of 98°C 15 seconds, 65°C 30 seconds, 72°C, 30 seconds, with data acquisition set to the 72°C extension. PCR1 reactions were cleaned up by adding of 4.5 μL of isopropanol, 54 μL of AMPure XP beads and incubation for 10 min. Beads were washed once with 80% ethanol, dried for 5 min, and eluted in 15 μl of water. Illumina flow cell adaptors were added by adding 15 μL 2X Phusion HF-PCR Master Mix and 0.4 μL P3solexa/P6solexa oligo mix and amplified: 98°C 2 min, 3 cycles of 98°C 15 seconds, 65°C 30 seconds, 72°C, 30s seconds. Final libraries were purified by addition of 48 μL of AMPure XP beads and incubation for 5 min. Beads were washed twice with 70% ethanol, dried for 5 min, and eluted in 20 μL of water. 1-2μL of libraries were quantitated by HS-DNA Bioanalyzer. Samples were deep sequenced on the Illumina NextSeq machine: single-end, no index, high-output, 75-bp cycle run.

Whole transcriptome RNA sequencing was performed using the methods described above for RNA extraction, library preparation and sequencing.

Nuclear/cytoplasmic RNA fractionation and sequencing

Fractionation was performed using the Cytoplasmic and Nuclear RNA Purification Kit (Cat. # 2100, Norgen Biotek Corp, Thorold, CA). Purified RNA was treated with DNase I followed by library preparation using the NuGEN V2 RNA-Sequencing Library Preparation kit (Tecan Genomics, Redwood City, CA). Both RNA and library quality were analyzed via a Bioanalyzer (Agilent, Santa Clara, CA). Sequencing was performed on a NextSeq 500 (Illumina, San Diego, CA): single-end, no index, high-output, 75-bp cycle run.

IF and RNAscope protocol

For immunofluorescence and RNAscope, infected NPCs were collected at 48 hr and plated onto 22- by 22-mm no. 1.5 coverslips. Cells were then fixed with 4% PFA in PBS for 15 minutes. For the RNAscope protocol, we followed manufacturer's instructions for adherent cell lines. Briefly, we first dehydrated the cells using 50%, 70% and then 100% ethanol in PBS. This was followed by a rehydration of the cells using 70% and then 50% of ethanol in PBS. Lastly, cells were fully rehydrated in PBS. Cells were then permeabilized by hydrogen peroxide, followed by protease 3 treatment. Next, we hybridized the RNAscope probes to the cells for 2 hours or O/N at 40C. Probe amplification was then performed, followed by labelling with Opal570 (Akoya Biosciences). Nuclei were stained using Hoechst 33258 (Thermofisher).

For immunofluorescence, cells were fixed, permeabilized by 0.1% Triton X-100 (Sigma Aldrich), blocked with 3% bovine serum albumin (Sigma Aldrich) in PBS. Cells were then immunostained with the indicated antibody, followed by the appropriate secondary. Lastly, nuclei were stained with Hoechst 33258.

Microscopy was performed on an LSM880 with Airyscan (Zeiss, Oberkochen, Germany) or an Olympus (Tokyo, Japan) FV3000RS. On the LSM880, imaging of Huh7-Lunet cells was performed with 20x magnification objective, while NPCs were imaged with a 63x oil objective. On the FV3000RS, the 20x objective was used for imaging NPCs. All images were taken as a Z-stack.

Western blot analysis

Cells were lysed in RIPA lysis buffer (50 mM Tris-HCl [pH 8], 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, supplemented with Halt protease inhibitor cocktail [Thermo Fisher Scientific]) to obtain whole-cell lysates or lysed using the NE-PER nuclear and cytoplasmic extraction kit (Thermo Fisher Scientific) to obtain cytoplasmic and nuclear fractions. Proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes (Bio-Rad, Hercules, CA). Proteins were visualized by chemiluminescent detection with ECL and visualized on a ChemiDoc MP Imaging System (Bio-Rad).

Computational and Statistical Analyses

For western blot analysis, differences in band intensity were quantified by densitometry using ImageJ (Schneider et al., 2012). Student's t-test was used for statistical analysis of western blots.

Imaris (Oxford Instruments, Abingdon, UK) was used for analysis of confocal images, using the surface function for polyA mRNA analysis and dot function identifying specific transcripts by RNA-scope. Nuclei were also bounded and identified by the surface function of Imaris. PolyA RNA-scope experiments were statistically analyzed using a linear mixed model to account for individual cell values across multiple biological replicates. Data are represented as means plus standard errors of the means (SEM). Gene set overlap statistics were performed using a hypergeometric test. Statistical significance was defined as follows: *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$. Biological replicates are defined as the same experimental design but performed sequentially, with a different cell passage number and on different days.

Analysis of RNA sequencing data

PCR duplicates were removed using unique molecular identifiers in the RT primer region. The adaptor and barcode sequences were trimmed and reads were mapped step-wise to viral (ZIKV), repetitive and finally non-repetitive (GRCh38) genomes. Bowtie2 indexes were generated using the 'bowtie2-build' command in Bowtie2 for the ZIKV (KU501215.1) RNA genome sequences. The specific parameters used for the FAST-iCLIP pipeline were as follows: -f 18 (trims 17 nt from the 5′ end of the read), -l 16 (includes all reads longer than 16 nt), –bm 29 (minimum MAPQ score from bowtie2 of 29 is required for mapping; unique mapping only), –tr 2,3 (repetitive genome) and –tn 2,3 (non-repetitive genome) RT stop intersection (n,m; where n = replicate number and m = number of unique RT stops required per n replicates). Using the –tr/tn 2,3 parameters, a minimum of six RT stops are required to support any single nucleotide identified as a crosslinking site.

Analysis of the sequencing data was performed using a custom analysis pipeline, with the peak finding software uploaded to Github (https://github.com/ChangLab/FAST-iCLIP/tree/lite). Other analyses were performed by aligning the reads to the human genome using STAR, followed by gene counts using Bedtools (Quinlan and Hall, 2010). Only reads in the canonical 3'UTR of the human transcripts were counted. The count distribution across the metagene of the CDS and 3'UTRs was visualized using deeptools (Ramírez et al., 2014). Log_2 fold changes were calculated by comparing RPKM values between Mock and ZIKV infected cells. Genes with fewer than 10 total read counts were excluded from analysis.

Heat maps and XY plots were produced using Prism 8 (GraphPad, San Diego, CA).

For total RNA and cytoplasmic/nuclear fractionated sequencing, reads were aligned to the human genome using the STAR aligner (Dobin et al., 2013) (version 2.7.5) followed by HTSeq (Anders et al., 2015) (version 0.12.3) to obtain counts and then using the DeSeq2 (Love et al., 2014) (version 1.28.1) pipeline to determine $\log_2$ fold changes in transcripts. Volcano plots were produced using the R-package, Enhanced Volcano (https://github.com/kevinblighe/EnhancedVolcano).

Acknowledgements

Stephen Floor, PhD and Michael Wilson, MD provided guidance and mentorship during this project as members of a thesis committee. Library preparation and QCs for sequencing was conducted by Mylinh Bernardi, BS at the Gladstone Genomics Core. The James B. Pendleton Charitable Trust supported the sequencing on the NextSeq 500. Meredith Calvert, PhD and Blaise Ndjamen, PhD of the Gladstone Histology and Light Microscopy core assisted with imaging and image analysis. Reuben Thomas, PhD in the Gladstone Bioinformatics core provided statistical analysis help. Wendy Runyon, BS in the Gladstone Stem Cell core helped with iPSC maintenance. John Carroll provided graphic design expertise, and Kathryn Claiborn, PhD provided manuscript editing.

Funding

This work was supported by the UCSF Medical Scientist Training Program (KEL), the UCSF Biomedical Sciences Program (KEL), the NINDS/NIH under award F31NS113432 (KEL), the NIAID/NIH under award 5R01AI097552-03 (KEL, MO), NIDA/NIH under award 5DP1DA038043-05 (KEL, MO), NIH/NIAID R01 AI141970 (JEC), NIH/NIAID F32AI112262 (PSS), NIH/NINDS R01 NS101996-01 (SF), NIH/NIAID U19AI1186101 to (NJK), DOD/DARPA HR0011-11-C- 0094 (PROPHECY, NJK), the Otellini Family Discovery Fellowship (KEL), Damon Runyon Cancer Research Foundation DRG-2286-17 (R.A.F.), and the James B. Pendleton Charitable Trust.

Author Contributions

KEL, MO contributed to experimental design, data acquisition, data analysis and manuscript writing. RAF, CB, JEC contributed to experimental design, data acquisition and data analysis. MMK, CS contributed to experimental design and data acquisition. KAF, TN, ST and GRK contributed to experimental design and data analysis. DJ-M, PSS and NJK contributed with computational analysis. MD, JK and SF contributed assistance with data acquisition.

Competing Interests

No competing or conflicts of interest to report.

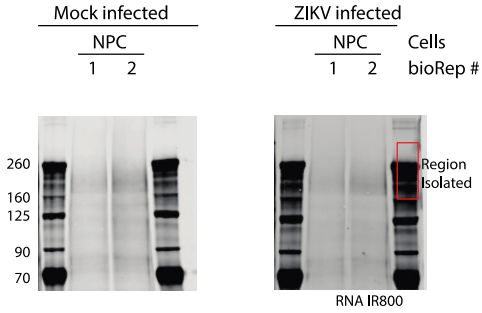
Sequencing Data

Raw Sequencing files were uploaded to GEO: To review GEO accession GSE160201: https://urldefense.proofpoint.com/v2/url?u=https-3A__www.ncbi.nlm.nih.gov_geo_query_acc.cgi-3Facc-3DGSE160201&d=DwIBAg&c=iORugZls2LlYyCAZRB3XLg&r=1hBFwiZQbvcsplNqQhwOzovvOfewuh5fldcgMxoR3Bw&m=Z4gCgqzdehtx01hOtBv7q1TkUQvLtq7LoaYKXZVxx28&s=SUUL5yZGVWmqLSEC93kiEvxDAOqBxKwUoNyTY_UDUI8&e=

Enter token irmxsyeevvinvuf into the box

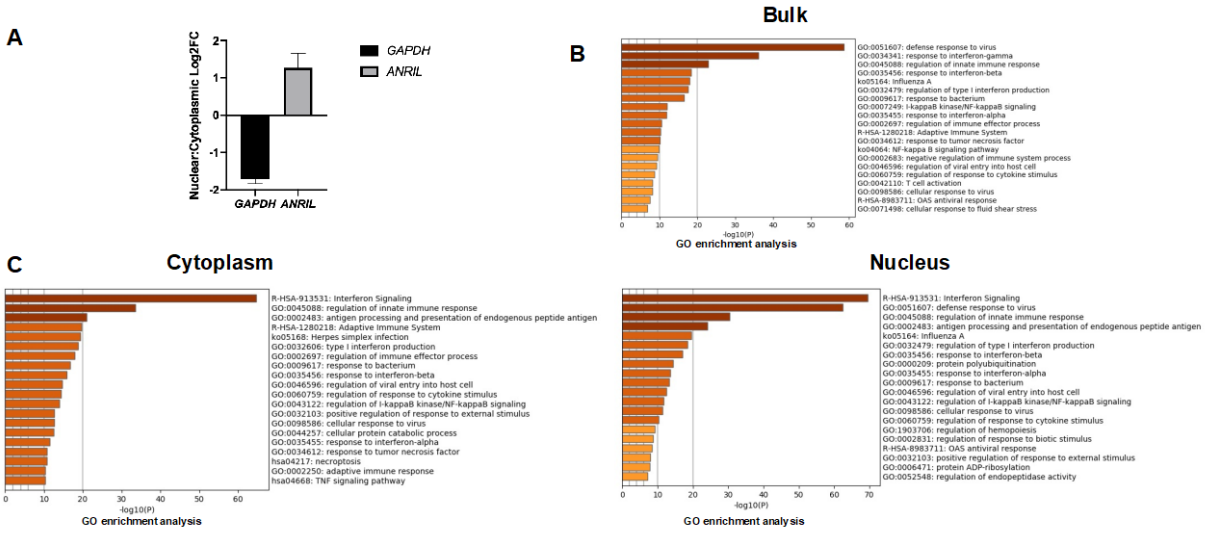
Supplemental Figures

Supplemental Figure 1



Supplemental Figure 1: Electrophoresis of RNA pulled down with UPF1, visualized using the IR handle ligated to the RNAs. The region bounded by the red box indicates the part of the gel excised and then analyzed by irCLIP and AP-MS. AP-MS analysis is indicated in Supplemental File 1.

Supplemental Figure 2



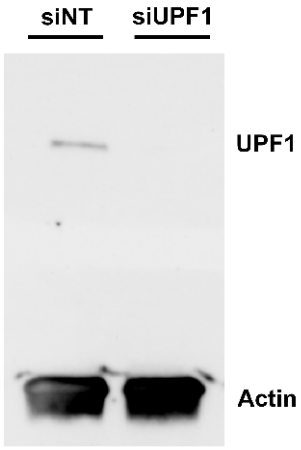
Supplemental Figure 2

A) Log_2 fold change from the fractionated RNA-sequencing between the nucleus and cytoplasm for markers of successful fractionation in the mock-infected samples: GAPDH for cytoplasm and ANRIL for nucleus.

B) Metascape Analysis of significantly upregulated transcripts found in the bulk RNA sequencing of Figure 2A. The top 400 upregulated genes were used to produce this GO clustering.

C) Metascape Analysis of significantly upregulated transcripts found in the cytoplasmic and nuclear fractionated RNA sequencing of Figure 2D. The top 400 and 300 upregulated genes were used to produce this GO clustering.

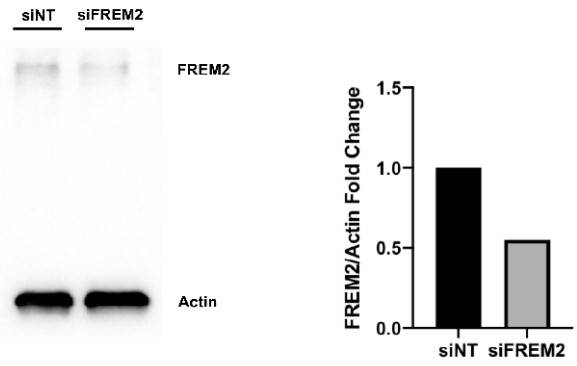
Supplemental Figure 3



Supplemental Figure 3

Western blot for UPF1 in siNT and siUPF1 treated cells. Actin is shown as a loading control.

Supplemental Figure 4



Supplemental Figure 4:

Western blot for FREM2 in siNT and siFREM2 treated NPCs after 7 days. Densitometric analyses of FREM2 were performed using ImageJ to quantify relative band intensities

Supplemental Files

Supplemental Files 1: Slice 1 and Slice 2

Results of Mass Spectrometry Analysis of slices extracted for irCLIP sequencing of ZIKV infected NPCs

Supplemental File 2:

Read counts from irCLIP of UPF1 and RPKM of Total RNA Sequencing of ZIKV-infected NPCs

Supplemental File 3:

DESeq2 Analysis of Total RNA Sequencing of Mock and ZIKV-infected NPCs

Supplemental File 4:

DESeq2 Analysis of Cytoplasmic RNA Sequencing of Mock and ZIKV-infected NPCs

Supplemental File 5:

DESeq2 Analysis of Nuclear RNA Sequencing of Mock and ZIKV-infected NPCs

References

Ajamian, L., Abel, K., Rao, S., Vyboh, K., Garcia-de-Gracia, F., Soto-Rifo, R., Kulozik, A.E., Gehring, N.H., and Mouland, A.J. (2015). HIV-1 Recruits UPF1 but Excludes UPF2 to Promote Nucleocytoplasmic Export of the Genomic RNA. Biomolecules *5*, 2808-2839.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Balistreri, G., Bognanni, C., and Mühlemann, O. (2017). Virus Escape and Manipulation of Cellular Nonsense-Mediated mRNA Decay. Viruses *9*.

Carmody, S.R., and Wente, S.R. (2009). mRNA nuclear export at a glance. J Cell Sci *122*, 1933-1937.

Chawla, R., Redon, S., Raftopoulou, C., Wischnewski, H., Gagos, S., and Azzalin, C.M. (2011). Human UPF1 interacts with TPP1 and telomerase and sustains telomere leading-strand replication. The EMBO journal *30*, 4047-4058.

Colak, D., Ji, S.-J., Porse, B.T., and Jaffrey, S.R. (2013). Regulation of axon guidance by compartmentalized nonsense-mediated mRNA decay. Cell *153*, 1252-1265.

Consortium, H.i. (2017). Developmental alterations in Huntington's disease neural cells and pharmacological rescue in cells and mice. Nat Neurosci *20*, 648-660.

Costa, F., Sarno, M., Khouri, R., de Paula Freitas, B., Siqueira, I., Ribeiro, G.S., Ribeiro, H.C., Campos, G.S., Alcântara, L.C., Reis, M.G.*, et al.* (2016). Emergence of Congenital Zika Syndrome: Viewpoint From the Front Lines. Ann Intern Med *164*, 689-691.

Cugola, F.R., Fernandes, I.R., Russo, F.B., Freitas, B.C., Dias, J.L.M., Guimarães, K.P., Benazzato, C., Almeida, N., Pignatari, G.C., Romero, S.*, et al.* (2016). The Brazilian Zika virus strain causes birth defects in experimental models. Nature *534*, 267-271.

de Paula Freitas, B., Ventura, C.V., Maia, M., and Belfort, R., Jr. (2017). Zika virus and the eye. Curr Opin Ophthalmol *28*, 595-599.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Fauci, A.S., and Morens, D.M. (2016). Zika Virus in the Americas--Yet Another Arbovirus Threat. The New England Journal of Medicine *374*, 601-604.

Feng, Q., Jagannathan, S., and Bradley, R.K. (2017). The RNA Surveillance Factor UPF1 Represses Myogenesis via Its E3 Ubiquitin Ligase Activity. Molecular Cell *67*, 23-251.e256.

Fiorini, F., Bagchi, D., Le Hir, H., and Croquette, V. (2015). Human Upf1 is a highly processive RNA helicase and translocase with RNP remodelling activities. Nat Commun *6*, 7581.

Fontaine, K.A., Leon, K.E., Khalid, M.M., Tomar, S., Jimenez-Morales, D., Dunlap, M., Kaye, J.A., Shah, P.S., Finkbeiner, S., Krogan, N.J.*, et al.* (2018). The Cellular NMD Pathway Restricts Zika Virus Infection and Is Targeted by the Viral Capsid Protein. mBio *9*.

Fortes, P., Beloso, A., and Ortín, J. (1994). Influenza virus NS1 protein inhibits pre-mRNA splicing and blocks mRNA nucleocytoplasmic transport. Embo j *13*, 704-712.

Gehring, N.H., Lamprinaki, S., Kulozik, A.E., and Hentze, M.W. (2009). Disassembly of exon junction complexes by PYM. Cell *137*, 536-548.

Gehring, N.H., Neu-Yilik, G., Schell, T., Hentze, M.W., and Kulozik, A.E. (2003). Y14 and hUpf3b Form an NMD-Activating Complex. Molecular Cell *11*, 939-949.

Göertz, G.P., Abbo, S.R., Fros, J.J., and Pijlman, G.P. (2018). Functional RNA during Zika virus infection. Virus Res *254*, 41-53.

Imamachi, N., Salam, K.A., Suzuki, Y., and Akimitsu, N. (2017). A GC-rich sequence feature in the 3′ UTR directs UPF1-dependent mRNA decay in mammalian cells. Genome Res *27*, 407-418.

Jaffrey, S.R., and Wilkinson, M.F. (2018). Nonsense-mediated RNA decay in the brain: emerging modulator of neural development and disease. Nat Rev Neurosci *19*, 715-728.

Kebaara, B.W., and Atkin, A.L. (2009). Long 3′-UTRs target wild-type mRNAs for nonsense-mediated mRNA decay in Saccharomyces cerevisiae. In Nucleic Acids Res, pp. 2771-2778.

Kim, Y.K., and Maquat, L.E. (2019). UPFront and center in RNA decay: UPF1 in nonsense-mediated mRNA decay and beyond. Rna *25*, 407-422.

Kumar, G.R., and Glaunsinger, B.A. (2010). Nuclear import of cytoplasmic poly(A) binding protein restricts gene expression via hyperadenylation and nuclear retention of mRNA. Mol Cell Biol *30*, 4996-5008.

Kurosaki, T., and Maquat, L.E. (2016). Nonsense-mediated mRNA decay in humans at a glance. J Cell Sci *129*, 461-467.

Kurosaki, T., Popp, M.W., and Maquat, L.E. (2019). Quality and quantity control of gene expression by nonsense-mediated mRNA decay. Nat Rev Mol Cell Biol *20*, 406-420.

Kuss, S.K., Mata, M.A., Zhang, L., and Fontoura, B.M. (2013). Nuclear imprisonment: viral strategies to arrest host mRNA nuclear export. Viruses *5*, 1824-1849.

Leon, K., and Ott, M. (2020). An 'Arms Race' between the Nonsense-mediated mRNA Decay Pathway and Viral Infections. Seminars in cell & developmental biology.

Li, C., Xu, D., Xu, Z., Ye, Q., Hong, S., Jiang, Y., Liu, X., Zhang, N., Shi, L., and Qin, C.-F. (2016). Zika Virus Disrupts Neural Progenitor Development and Leads to Microcephaly in Mice. Cell Stem Cell *19*, 120-126.

Li, M., Johnson, J.R., Truong, B., Kim, G., Weinbren, N., Dittmar, M., Shah, P.S., Von Dollen, J., Newton, B.W., Jang, G.M.*, et al.* (2019). Identification of antiviral roles for the exon-junction complex and nonsense-mediated decay in flaviviral infection. Nat Microbiol *4*, 985-995.

Link, N., Chung, H., Jolly, A., Withers, M., Tepe, B., Arenkiel, B.R., Shah, P.S., Krogan, N.J., Aydin, H., Geckinli, B.B.*, et al.* (2019). Mutations in ANKLE2, a ZIKA Virus Target, Disrupt an Asymmetric Cell Division Pathway in Drosophila Neuroblasts to Cause Microcephaly. Dev Cell *51*, 713-729.e716.

Liu, L., Chen, Z., Zhang, X., Li, S., Hui, Y., Feng, H., Du, Y., Jin, G., Zhou, X., and Zhang, X. (2019). Protection of ZIKV infection-induced neuropathy by abrogation of acute antiviral response in human neural progenitors. Cell Death Differ *26*, 2607-2621.

Loh, B., Jonas, S., and Izaurralde, E. (2013). The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. Genes Dev *27*, 2125-2138.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Lykke-Andersen, S., Chen, Y., Ardal, B.R., Lilje, B., Waage, J., Sandelin, A., and Jensen, T.H. (2014). Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. Genes Dev *28*, 2498-2517.

Martínez-Cerdeño, V., and Noctor, S.C. (2018). Neural Progenitor Cell Terminology. Front Neuroanat *12*, 104.

Mohd Ropidi, M.I., Khazali, A.S., Nor Rashid, N., and Yusof, R. (2020). Endoplasmic reticulum: a focal point of Zika virus infection. J Biomed Sci *27*, 27.

Musso, D., and Gubler, D.J. (2016). Zika Virus. Clinical Microbiology Reviews *29*, 487-524.

Ojha, C.R., Rodriguez, M., Lapierre, J., Muthu Karuppan, M.K., Branscome, H., Kashanchi, F., and El-Hage, N. (2018). Complementary Mechanisms Potentially Involved in the Pathology of Zika Virus. Front Immunol *9*, 2340-2340.

Oliveira, E.R.A., Mohana-Borges, R., de Alencastro, R.B., and Horta, B.A.C. (2017). The flavivirus capsid protein: Structure, function and perspectives towards drug design. Virus Res *227*, 115-123.

Pavlakis, E., Chiotaki, R., and Chalepakis, G. (2011). The role of Fras1/Frem proteins in the structure and function of basement membrane. Int J Biochem Cell Biol *43*, 487-495.

Peccarelli, M., and Kebaara, B.W. (2014). Regulation of natural mRNAs by the nonsense-mediated mRNA decay pathway. Eukaryot Cell *13*, 1126-1135.

Pierson, T.C., and Diamond, M.S. (2018). The emergence of Zika virus and its new clinical syndromes. Nature *560*, 573-581.

Popp, M.W., and Maquat, L.E. (2013). Organizing principles of mammalian nonsense-mediated mRNA decay. Annu Rev Genet *47*, 139-165.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res *42*, W187-191.

Roth, H., Magg, V., Uch, F., Mutz, P., Klein, P., Haneke, K., Lohmann, V., Bartenschlager, R., Fackler, O.T., Locker, N.*, et al.* (2017). Flavivirus Infection Uncouples Translation Suppression from Cellular Stress Responses. mBio *8*.

Saade, M., Ferrero, D.S., Blanco-Ameijeiras, J., Gonzalez-Gobartt, E., Flores-Mendez, M., Ruiz-Arroyo, V.M., Martínez-Sáez, E., Ramón, Y.C.S., Akizu, N., Verdaguer, N.*, et al.* (2020). Multimerization of Zika Virus-NS5 Causes Ciliopathy and Forces Premature Neurogenesis. Cell Stem Cell *27*, 920-936.e928.

Sakuma, T., Davila, J.I., Malcolm, J.A., Kocher, J.P., Tonne, J.M., and Ikeda, Y. (2014). Murine leukemia virus uses NXF1 for nuclear export of spliced and unspliced viral transcripts. J Virol *88*, 4069-4082.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature Methods *9*, 671-675.

Schuessler, A., Funk, A., Lazear, H.M., Cooper, D.A., Torres, S., Daffis, S., Jha, B.K., Kumagai, Y., Takeuchi, O., Hertzog, P.*, et al.* (2012). West Nile virus noncoding subgenomic RNA contributes to viral evasion of the type I interferon-mediated antiviral response. J Virol *86*, 5708-5718.

Shah, P.S., Link, N., Jang, G.M., Sharp, P.P., Zhu, T., Swaney, D.L., Johnson, J.R., Von Dollen, J., Ramage, H.R., Satkamp, L.*, et al.* (2018). Comparative Flavivirus-Host Protein Interaction Mapping Reveals Mechanisms of Dengue and Zika Virus Pathogenesis. Cell *175*, 1931-1945.e1918.

Shao, Q., Herrlinger, S., Yang, S.-L., Lai, F., Moore, J.M., Brindley, M.A., and Chen, J.-F. (2016). Zika virus infection disrupts neurovascular development and results in postnatal microcephaly with brain damage. Development (Cambridge, England) *143*, 4127-4136.

Singh, A.K., Choudhury, S.R., De, S., Zhang, J., Kissane, S., Dwivedi, V., Ramanathan, P., Petric, M., Orsini, L., Hebenstreit, D.*, et al.* (2019). The RNA helicase UPF1 associates with mRNAs co-transcriptionally and is required for the release of mRNAs from gene loci. Elife *8*.

Slavotinek, A.M., and Tifft, C.J. (2002). Fraser syndrome and cryptophthalmos: review of the diagnostic criteria and evidence for phenotypic modules in complex malformation syndromes. J Med Genet *39*, 623-633.

Souza, B.S.F., Sampaio, G.L.A., Pereira, C.S., Campos, G.S., Sardi, S.I., Freitas, L.A.R., Figueira, C.P., Paredes, B.D., Nonaka, C.K.V., Azevedo, C.M.*, et al.* (2016).

Zika virus infection induces mitosis abnormalities and apoptotic cell death of human neural progenitor cells. Scientific reports *6*, 39775.

Stoeckli, E.T. (2018). Understanding axon guidance: are we nearly there yet? Development *145*.

Tang, H., Hammack, C., Ogden, Sarah C., Wen, Z., Qian, X., Li, Y., Yao, B., Shin, J., Zhang, F., Lee, Emily M.*, et al.* (2016). Zika Virus Infects Human Cortical Neural Progenitors and Attenuates Their Growth. Cell Stem Cell *18*, 587-590.

Timmer, J.R., Mak, T.W., Manova, K., Anderson, K.V., and Niswander, L. (2005). Tissue morphogenesis and vascular stability require the Frem2 protein, product of the mouse myelencephalic blebs gene. Proc Natl Acad Sci U S A *102*, 11746-11750.

Toma, K.G., Rebbapragada, I., Durand, S., and Lykke-Andersen, J. (2015). Identification of elements in human long 3' UTRs that inhibit nonsense-mediated decay. Rna *21*, 887-897.

Varsally, W., and Brogna, S. (2012). UPF1 involvement in nuclear functions. Biochem Soc Trans *40*, 778-783.

Ventura, C.V., and Ventura, L.O. (2018). Ophthalmologic Manifestations Associated With Zika Virus Infection. Pediatrics *141*, S161-s166.

Vicenti, I., Boccuto, A., Giannini, A., Dragoni, F., Saladini, F., and Zazzi, M. (2018). Comparative analysis of different cell systems for Zika virus (ZIKV) propagation and evaluation of anti-ZIKV compounds in vitro. Virus Res *244*, 64-70.

Wen, F., Armstrong, N., Hou, W., Cruz-Cosme, R., Obwolo, L.A., Ishizuka, K., Ullah, H., Luo, M.H., Sawa, A., and Tang, Q. (2019). Zika virus increases mind bomb 1 levels, causing degradation of pericentriolar material 1 (PCM1) and dispersion of PCM1-containing granules from the centrosome. J Biol Chem *294*, 18742-18755.

Zarnegar, B.J., Flynn, R.A., Shen, Y., Do, B.T., Chang, H.Y., and Khavari, P.A. (2016). irCLIP platform for efficient characterization of protein-RNA interactions. Nat Methods *13*, 489-492.

Chapter 6

General Discussion and Future Perspectives

General discussion

Outbreaks and pandemics cause considerable health burdens in the 21st century. Although SARS-CoV2 and ZIKV have similar replication patterns, they transmit and infect differently. SARS-CoV2 spreads between humans through respiratory fluids, jumped from animal reservoir to human, spread globally, and infected more than 182 million, causing COVID-19 disease. Even though SARS-CoV2 has a lower mortality rate than SARS-CoV and MERS-CoV, the ongoing COVID-19 pandemic has already taken ~ 4 million lives due to the global spread. From the beginning of the pandemic, WHO and CDC issued control measures like masks, social distancing, proper hygiene, and lockdown to limit the spread of the virus to communities and other countries. Lockdown and social distancing primarily restrict human movement and interaction, thus limit the chances of viral spread in a population. Mask and proper hygiene reduce the chances of spreading or getting infected by SARS-CoV2. SARS-CoV2 spread globally even after preventive measures were taken by many countries. In addition, the SARS-CoV2 infections can be asymptomatic and have more prolonged incubation which allows the virus to spread efficiently without knowledge of the infected patient. Massive testing and reliable, easy-to-access tests can identify the percentage of infected people in a given population. Therefore, public awareness and support, along with government initiatives, are crucial to fight this pandemic.

Countries that followed strict mask rules and lockdown saw reduced COVID-19 mortality in the first and second waves of the pandemic. Currently, there are several approved vaccines available. These vaccines came in a record amount of time, and their efficacy is remarkable (1, 2). The Food and Drug Administration (FDA) approved two mRNA vaccines (Moderna and Pfizer) that showed almost 95% protection against SARS-CoV2 infection (3). This achievement is remarkable for the field of vaccines and infectious diseases. It is the first time any mRNA vaccine was approved and administered in such a large population. The FDA also approved the Janssen COVID-19 vaccine, an adenovirus vector-based vaccine, for emergency use (4-7). Other available vaccines (AstraZeneca, Sputnik V) also showed promising results (8). One shot of Pfizer or AstraZeneca can reduce transmission by half (9). When a specific proportion of a population became immune to a particular infectious disease, it provides indirect protection against that disease to an unimmunized person. This indirect protection is called herd immunity, which varies depending infection rate of each infectious disease. A large number of the population needs to be vaccinated to achieve herd immunity for COVID-19 (10). The more people get the COVID-19 vaccine, the more protection against COVID-19 will be. Awareness of getting vaccinated is paramount to stop the waves of any viral pandemic.

At the same time, the emerging spike mutant variants raise public concerns and the efficiency of vaccines (10). Studying the functions of the spike proteins and variants will be beneficial in understanding the dynamics of the SARS-CoV2 pandemic. Understanding which spike mutation or variant can cause increased infectivity or evade the immune system will help to design a future successful vaccine protecting against variants. At the same time, exploring other SARS-CoV2 proteins will be helpful to understand the pathogenesis of the COVID-19. Currently, COVID-19 has no cure. The focus is to reduce the severity and mortality caused by SARS-CoV2. Studying the similarity and dissimilarity of SARS-CoV2 proteins' function allows finding shared and unique pathways and mechanisms to target SARS-CoV2 that might save lives by vaccination, treatment, and other control measures.

Even though the 2015 ZIKV epidemic has ended, vector-borne infections are still threats for new outbreaks. Every year, many people die in DENV infection-related deaths. Mosquitoes that live close to humans can spread many viral and other infectious diseases. In addition, human activities and deforestation are causing displacement of mosquitoes' habitats and allowing humans to get exposed to them. Thus, new or emerging arboviral or flaviviral outbreaks can happen. These outbreaks allow viral evolution to transmit better in human and poses a further threat of being circulated continuously. Even though ZIKV was discovered so long ago, up until the recent epidemic, it was considered a mild infection. There is still no specific treatment or approved vaccine for ZIKV is available. The affected areas usually experience or have high chances of DENV outbreaks too. The temperate zone is getting widened with global warming, increasing places where mosquitoes can live. Urbanization with improper plastic waste disposal allows more breeding grounds for mosquitoes. Mosquitoes can quickly spread viral and other mosquito-borne diseases by living close to human habitats, mostly in dense living settings. The mosquito-borne virus, like ZIKV, needs to adapt and evolve in both humans and vectors. ZIKV targets, hijacks, and interacts with host-vector proteins and pathways for adaptation in host and vector. Thus these viruses interact and target lots of host proteins for survival, transmission, and efficient replication. If we understand how these viral proteins interact with host proteins, we can find a cure/vaccine and prevent viral pathology. At the same time, understanding pathways targeted by the virus will enrich our knowledge about viral replication and evolution.

In **Chapter 1**, a brief history of past and recent outbreaks and pandemics was initially discussed to show how the infectious disease devastated human life repeatedly. RNA viruses surprisingly caused all major outbreaks, epidemics, and pandemics in the last 20 years. Among these viruses, three different Coronoviruses pose a global threat, and SARS-CoV2 spread

globally. On the other hand, flaviviral outbreaks are not new, but ZIKV posed a distinct and unique danger globally. In this chapter, the general characteristics of SARS-CoV2 and ZIKV are described. Both viruses share similarities and dissimilarities. Although they have a similar replication cycle, they have different viral proteins with different functions. Different ways of spillover events (zoonosis vs. vector-borne) happened in both viruses. Three viral proteins functions and characteristics are discussed to understand how both viruses replicate, spread, and cause pathogenesis. SARS-CoV2 spike and ORF8 are significant for SARS-CoV2 transmission and viral fitness. The discussion focused on the mutations and functions of both proteins. Spike protein is essential for entry and viral fusion. Mutations that allow better entry or antibody escape allow the virus to spread faster and efficiently.

On the other hand, the function of SARS-CoV2 ORF8 is mainly unknown. Several independent studies show that ORF8 is important for immune evasion and pathogenesis. ORF8 targets the IL17RA pathway involved in cellular immune defense against viruses, bacteria, and fungus. IL17RA is linked to virus-mediated lung injury. SARS-CoV2 pathogenesis seen in COVID-19 probably due to dysregulation of IL17RA during viral infection. More research is needed to understand why ORF8 targets this pathway and the role of this pathway in SARS-CoV2 pathogenesis.

Another viral protein is ZIKV capsid, which is necessary for its replication cycle and viral assembly. Capsid interacts with many host proteins; it targets host cellular pathways to enhance viral replication. Although ZIKV replicates in the cytoplasm, ZIKV capsid localizes in the nuclear environment. ZIKV capsid targets an evolutionarily conserved NMD pathway. The nuclear role of the ZIKV capsid is mainly unknown. One way to understand virus-mediated pathogenesis is to study host-virus protein-protein interactions and targeted pathways by the virus. For both viruses, studying host-virus interactions helps to understand the diseases and finding novel mechanisms that will help to find cures or medications.

Chapter 2 identifies a new SARS-CoV2 variant(Epsilon) that emerged in California, USA—using whole-genome sequencing from samples collected from 44 counties in California. This variant harbors three unique spike mutations S13I, W152C, and L452R. We found up to 20% increases in transmission compared to the wild-type variant. Furthermore, L452R mutation has increased transmissibility in the cell line and organoid model. The increased number of infected patients with this lineage suggests higher transmissibility than wild type. We also found that this variant has some resistance to antibodies from convalescent (4-6.7 fold decrease) and vaccine

(2 fold decrease) patients. This study describes infectivity, transmission, and antibody neutralization of Epsilon variant (B.1.427 and B.4.429 lineages) (11).

Chapter 3 focuses on identifying and validating the conserved host factors targeted by three coronaviruses (SARS-CoV, SARS-CoV2, and MERS-CoV). We did an immunofluorescence localization study to finds cellular localization of all these viral proteins during overexpression in HelaM cells. We used comparative proteomics and functional genetics to identify unique or conserved host factors to all three viruses. We also generated antibodies against many SARS-CoV2 proteins. We found that among these viruses, similar proteins have similar localization. We conclude that localization is unlikely the cause of different host targeting mechanisms seen in these viruses. The host-virus protein-protein interactions and genetic perturbations are more relevant to find shared and unique targeting mechanisms. Functional studies, structural analysis, and real-world scenario data can validate shared host factors from comparative protein-protein interaction comparison. To do that, we performed RNAi in A549-ACE2 and CRISPR-KO in Caco2 cells and identified several targets that are unique for SARS-CoV2 and shared targets among these three viruses. We then validated the functions, interactions, and clinical relevancy of some of these factors Tom70, IL17RA, PGES-2, and SigmaR1. Cryo-EM microscopy, immunoprecipitation, and confocal microscopy showed that ORF9b interacts with Tom70, co-localizes together, and ORF9b binds Tom70 at the substrate-binding site. Tom70 is involved in mitochondrial antiviral signaling. We also found SARS-CoV2 infection decreased Tom-70 expression. Then we showed, SARS-CoV2 ORF8 uniquely functionally interacts with IL17RA and is confirmed by immunoprecipitation. IL17A is a crucial member of IL17RA signaling did not disrupt this interaction. Knockdown of IL17RA decreased SARS-CoV2 replication in A549-ACE2 cells. Less hospitalization of COVID-19 patients was associated with higher sIL17RA(soluble IL17RA). We also found that IL17A treatment reduced SARS-CoV2 replication in A549-ACE2 cells during infection. We then compared data from patients who received or did not receive drugs that target host factors found on the screen. We showed that drugs that target or inhibit PGES-2 and SIGMR1 were effective against COVID-19 severity (12). This study shows that repurposing drugs based on host-viral protein interaction and functional genetic screens along with real-world data thus can be applied.

In **chapter 4**, we discover the role of ZIKV capsid targeting host NMD pathway for enhancing viral replication. ZIKV infection in Huh7 cells and NPCs caused elevated NMD transcripts. This result suggests ZIKV infection disrupts the NMD pathways. The NMD is an RNA surveillance pathway that targets RNA viruses. UPF1 an essential protein for the NMD pathway.

We compared upregulated genes between ZIKV infection and UPF1 knockdown; found a significant overlap between them. Analyzing the ZIKV host-protein AP-MS interactome, we also found several members of the NMD pathway interact with ZIKV capsid. By immunoprecipitation, we showed that ZIKV capsid interacts with UPF1 and UPF1 mutants. ZIKV capsid expression and viral infection decreased UPF1 expression. We also found that ZIKV capsid causes proteasomal-dependent degradation of nuclear UPF1. Knockdown of UPF1 caused increased ZIKV replication (13). From these observations, we concluded that ZIKV capsid targets nuclear UPF1 to enhance ZIKV replication.

In **chapter 5**, we further investigated ZIKV-mediated UPF1 downregulation. Using infrared crosslinking immunoprecipitation and RNA sequencing (irCLIP-Seq) in ZIKV infected NPCs, we saw a decrease in UPF1-host transcripts interaction. In addition, our data suggest that in the cytoplasm, degradation of UPF1-occupied transcripts occurred during ZIKV infection. We also saw accumulation transcripts in the nucleus. Thus, upon ZIKV infection or capsid expression results in host transcripts accumulation in the nucleus.

Furthermore, we found that the knockdown of UPF1 leads to *FREM2* mRNA retention in the nucleus, thus causing reduced FREM2 protein expression. FREM2 protein is essential for differentiation during neurogenesis (14). This retention alters the expression of differentiation markers in NPCs. Thus, ZIKV infection causes decreased UPF1 expression, leading to premature differentiation via *FREM2* mRNA retention. Therefore this suggests that UPF1 downregulation causes altered differentiation potential of neurogenesis during ZIKV infection.

To prevent or cure a viral infection, understanding the molecular mechanism of pathogenesis is essential. In this regard, studying individual viral proteins is a good starting point. Viral mutations that allow the virus to replicate or infect efficiently cause higher spread to the population and become a dominant variant. Viral proteins interact with multiple proteins and have numerous functions. Viral proteins evolved to inhibit immune responses from the host to enhance replication. Viruses hijack host cellular machinery, disrupts pathways, down or upregulate certain protein functions by interacting with many host proteins. Mutations in the viral proteins help to evade the host immune system or defense mechanism. At the same time, viruses acquire mutations to binds these or other host proteins to disrupt their normal function, thus facilitating increased viral mRNA translation and infectivity. Disruption of normal processes or function in cells can result in devastating effects. The pathogenesis differs from virus to virus, depending on which cell, organ, and age group the virus targets. Thus understanding the molecular mechanism behind the pathogenesis of each virus is crucial to prevent or cure a viral disease.

Future Perspectives

New SARS-COV2 variants are emerging in different areas of the world. Spike D614G mutations allowed increased transmissibility, and all highly circulated variants have this mutation. We identified one new SARS-CoV2 variant in California, the USA, which has three Spike mutations (S13I, W152C, and L452R) alongside D614G (11). Other new variants also have L452R mutations. One variant which does not have D614G also has L452R mutation. (15). It would be interesting to see the infectivity and transmission of L452R spike mutation with or without D614G mutation. We also reported a hydrophobic patch near L452R, at F490 and L492 positions. Currently, the role of this hydrophobic patch is unknown. SARS-CoV2 ORF8 is a mysterious protein with immune evasion properties. ORF8 role in SARS-CoV2 species adaptation needs to be studied. At the same time, several ORF8 mutations and deletions emerged (16, 17). Effects of these mutations and deletions probably have different effects on viral replication and fitness. Whether ORF8 enhances or reduces viral transmission or replication is still unknown. On the other hand, we found that ORF8 interacts with IL17RA. IL17A treatment did not disrupt the interactions in the cell line (12). It would be interesting to see whether ORF8 can mimic the functions of IL17A. At the same time, it is essential to see whether SARS-CoV (ORF8a and ORF8b) and MERS-CoV (ORF8b) target the IL17RA pathway. ORF8 also interacts with other host proteins, and many of them are metalloproteases (19, 20). The effect of ORF8 and the metalloprotease interaction are unknown. In cell culture and patient serum, ORF8 is secreted, but which proteins ORF8 interacts with secreted forms are unknown (21). Whether ORF8 also affects cellular secretion is currently unknown. The immune evasion properties of ORF8 are pretty interesting (22-27). COVID-19 patients produce antibodies against ORF8; whether it is neutralizing or binding is unknown. Whether it plays more role in also adaptive immunity is worth investigating. At the same time, the inclusion of the ORF8 antibody in serological testing can be beneficial. Understanding the molecular biology of SARS-CoV2 spike and ORF8 will help to understand transmission and pathogenesis.

Flaviviral outbreaks are continuing problems in many parts of the world. ZIKV epidemic caught global attention in 2015 by causing increased fetal microcephaly cases. ZIKV capsid is a multi-functional structural protein with diverse roles. Here we showed how ZIKV targets the NMD pathway and downregulates the master regulator, UPF1, which leads to host *FREM2* mRNA accumulation in the nucleus (13, 14). It will be interesting to overexpress UPF1 and FREM2 and see the effect on ZIKV replication and pathogenesis. Besides this, multiple host nuclear proteins are interacting with ZIKV capsid. ZIKV capsid also binds all four nucleotides, but it is unknown

whether it happens during ZIKV infection and capsid overexpression. There are some chromatin remodelers found in the ZIKV interactome (28). Chromatin remodelers have a prominent role in development. Whether ZIKV capsid disrupts, chromatin remodelers' function is still unknown. If a specific role exists, using an activator or inhibitor of that chromatin remodeler can lessen the consequences of ZIKV infection. Capsid inhibitors can also be a therapeutic option. Targetting capsid probably will reverse capsid-mediated disruption of host cellular process, thus ZIKV pathogenesis.

This work studied two RNA viruses (SARS-CoV2 and ZIKV) and their molecular mechanisms of pathogenesis. First, we identified a new variant (Epsilon) of SARS-CoV2 in California. Then we looked at how this variant transmits and evades the immune system. We also aimed to find shared targets by three coronaviruses. At the same time, we identified unique targets of SARS-CoV2 and validated their role both in vitro and patient data. Then we showed ZIKV targets host RNA surveillance pathway where a viral protein disrupts and downregulate the pathway. Disruption of this pathway leads to developmental defects. Thus, understanding viral pathogenesis in the context of host-virus interactions can help to understand their molecular mechanism and pathogenesis. Here, we discussed a total of three viral proteins from two viruses. Other viral proteins also need investigations to find their role in viral infection, transmission, and pathogenesis. Understanding the molecular mechanism of viral infection, transmission, and pathogenesis will help us find a better cure or intervention. It can also prepare us better for future pandemics, epidemics, and outbreaks.

References

1. *COVID research: a year of scientific milestones.* Nature, 2021.
2. Fauci, A.S., *The story behind COVID-19 vaccines.* Science, 2021. **372**(6538): p. 109.
3. Tenforde, M.W., et al., *Effectiveness of Pfizer-BioNTech and Moderna Vaccines Against COVID-19 Among Hospitalized Adults Aged ≥65 Years - United States, January-March 2021.* MMWR Morb Mortal Wkly Rep, 2021. **70**(18): p. 674-679.
4. Mullard, A., *FDA authorizes first single-shot COVID-19 vaccine.* Nat Rev Drug Discov, 2021. **20**(4): p. 251.
5. FDA. *COVID-19 Vaccines*. 2021 [cited 2021; Available from: https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-vaccines.
6. Release, F.N. *Coronavirus (COVID-19) Update: FDA Authorizes Pfizer-BioNTech COVID-19 Vaccine for Emergency Use in Adolescents in Another Important Action in Fight Against Pandemic*. 2021; Available from: https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-pfizer-biontech-covid-19-vaccine-emergency-use.
7. FDA. *Moderna COVID-19 Vaccine*. 2021; Available from: https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/moderna-covid-19-vaccine.
8. Rogliani, P., et al., *SARS-CoV-2 Neutralizing Antibodies: A Network Meta-Analysis across Vaccines.* Vaccines (Basel), 2021. **9**(3).
9. Mahase, E., *Covid-19: One dose of vaccine cuts risk of passing on infection by as much as 50%, research shows.* BMJ, 2021. **373**: p. n1112.
10. Aschwanden, C., *Five reasons why COVID herd immunity is probably impossible.* Nature, 2021. **591**(7851): p. 520-522.
11. Deng, X., et al., *Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant.* Cell, 2021.
12. Gordon, D.E., et al., *Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms.* Science, 2020. **370**(6521).
13. Fontaine, K.A., et al., *The Cellular NMD Pathway Restricts Zika Virus Infection and Is Targeted by the Viral Capsid Protein.* mBio, 2018. **9**(6).
14. Leon, K., et al., *Zika Virus Infection Prevents Host mRNA Nuclear Export by Disrupting UPF1 Function.* bioRxiv, 2020: p. 2020.12.03.410837.

15. Augustin Anoh, E., et al., *SARS-CoV-2 variants of concern, variants of interest and lineage A.27 are on the rise in Côte d'Ivoire.* medRxiv, 2021: p. 2021.05.06.21256282.
16. Pereira, F., *SARS-CoV-2 variants combining spike mutations and the absence of ORF8 may be more transmissible and require close monitoring.* Biochem Biophys Res Commun, 2021. **550**: p. 8-14.
17. Alkhansa, A., G. Lakkis, and L. El Zein, *Mutational analysis of SARS-CoV-2 ORF8 during six months of COVID-19 pandemic.* Gene Rep, 2021. **23**: p. 101024.
18. Young, B.E., et al., *Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study.* Lancet, 2020. **396**(10251): p. 603-611.
19. Gordon, D.E., et al., *A SARS-CoV-2 protein interaction map reveals targets for drug repurposing.* Nature, 2020. **583**(7816): p. 459-468.
20. Chasapis, C.T., et al., *A SARS-CoV-2 -human metalloproteome interaction map.* J Inorg Biochem, 2021. **219**: p. 111423.
21. Wang, X., et al., *Accurate Diagnosis of COVID-19 by a Novel Immunogenic Secreted SARS-CoV-2 orf8 Protein.* mBio, 2020. **11**(5).
22. Hassan, S.S., et al., *A unique view of SARS-CoV-2 through the lens of ORF8 protein.* Comput Biol Med, 2021. **133**: p. 104380.
23. Flower, T.G., et al., *Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein.* Proc Natl Acad Sci U S A, 2021. **118**(2).
24. Meinberger, D., et al., *Analysis of IgM, IgA, and IgG isotype antibodies Directed against SARS-CoV-2 spike glycoprotein and ORF8 in the course of COVID-19.* Sci Rep, 2021. **11**(1): p. 8920.
25. Zhang, Y., et al., *The ORF8 protein of SARS-CoV-2 mediates immune evasion through down-regulating MHC-Iota.* Proc Natl Acad Sci U S A, 2021. **118**(23).
26. Park, M.D., *Immune evasion via SARS-CoV-2 ORF8 protein?* Nat Rev Immunol, 2020. **20**(7): p. 408.
27. Li, J.Y., et al., *The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway.* Virus Res, 2020. **286**: p. 198074.
28. Shah, P.S., et al., *Comparative Flavivirus-Host Protein Interaction Mapping Reveals Mechanisms of Dengue and Zika Virus Pathogenesis.* Cell, 2018. **175**(7): p. 1931-1945 e18.

Summary

Summary

Viral epidemics and pandemics cause death tolls and severe complications. Emerging viral pandemics and outbreaks are linked to spillover events resulting from human encroachment, deforestation, animal habitat destruction, and global warming. These activities lead to unexpected and accidental exposure to wild animals. A natural host or vector of a virus thus has a chance to spread to humans and gain mutations allowing further spread. At the same time, densely crowded transportations increases the possibility of an outbreak spreading to other countries or distant parts of a country. Growing human population and dense urban living conditions pose more threats for a new or emerging infectious disease to cause outbreaks. The current SARS-CoV2 pandemic and past ZIKV epidemic are both results of spillover events. Without any pre-existing immunity, these viruses can cause higher co-morbidity and mortality. With more infections in a population, these viruses will acquire more mutations to become more adaptable, persistent, and transmissible, thus increasing viral fitness. Viral proteins interact with host proteins to benefit viral replication and infection, leading to higher viral transmission and persistent infection in the host. The host-viral interactions for fitness can explain viral evolution and pathogenesis. At the same time, these interactions can usually disrupt normal function or induction of pathways leading to altered expression or disrupted function of specific protein or pathway. The altered and dysregulated pathway might benefit the virus but will have devastating effects on the host, which appear as symptoms or complications. Studying the role of viral proteins in pathogenesis can help to find a cure or better preventive measures. A starting point is to understand these roles by looking at individual viral proteins and how they interact with host proteins. Finding out how these interactions lead to dysregulation of a hijacked pathway gives insight into the molecular mechanism of viral pathogenesis. This thesis investigated the role of viral proteins of SARS-CoV2 and ZIKV on viral transmission and pathogenesis.

In **Chapter 1**, the history, transmission route, and molecular biology of SARS-CoV2 and ZIKV were discussed. Both of these viruses contain an RNA genome with a similar viral life cycle. However, their transmission routes and animal reservoirs are entirely different. The target age group and severity of diseases by these viruses are also different. SARS-CoV2 originated in bat and spillover to humans, causing the current COVID-19 pandemic. SARS-CoV2 spike protein is a structural protein subjected to many mutations and essential for viral entry to susceptible cells. Mutations accumulating in spike protein can change the transmission and infectivity of the virus and thus, give rise to many variants. These variants may be able to infect cells more efficiently than the original virus. Several spike mutations seem to be responsible for this. At the same time,

there are specific mutations in spike protein that prevent our immune system from recognizing the virus properly. For example, the L452R mutation in the Epsilon SARS-CoV2 variant appears to be responsible for increased viral transmission, infectivity, and immune system evasion.

ORF8 is an accessory protein of SARS-CoV2, which is linked to immune evasion. On the other hand, the IL17RA pathway is dysregulated in several viral infections. Interestingly, in COVID-19 patients, elevated end products (cytokines) of the IL17RA pathway have been found. Proteomics analysis and in-vitro experiment show that SARS-CoV2 ORF8 protein interacts with IL17RA. It has also been found that ORF8 can dysregulate the IL17RA pathway. The spike protein and ORF8 play an important role in viral transmission and pathogenesis, but most of their biology is yet to be discovered.

In 2015, the ZIKV epidemic caused increased fetal microcephaly in babies. The epidemic ended in 2016. Fetal microcephaly, a neurodevelopmental disorder, is characterized by the smaller head size of the fetus. ZIKV capsid is a structural protein that plays a crucial role in viral life cycles. Furthermore, capsid also interacts and targets many host pathways and contributes to neurodevelopmental defects. ZIKV capsid also localizes to the nuclear environment and targets crucial pathways of the host, like the NMD pathway. NMD pathway is an RNA surveillance pathway that targets RNA viruses and has a significant role in development, specifically neurodevelopment. ZIKV capsid interacts with UPF1, an essential member of the NMD pathway, and reduces its expression. This reduction of expression helps viral replication. Furthermore, the reduced expression of UPF1 causes the accumulation of *FREM2* mRNA in the nucleus, thus reduce FREM2 protein expression. FREM protein is essential for neurodevelopment. Reduced expression of FREM2 protein changes the differentiation pattern of NPC. Thus capsid mediated targeting of the NMD pathway caused ZIKV pathogenesis.

Host-viral interactions reveal cellular targets of the virus. Understanding how these interactions disrupt or dysregulate pathways give us molecular insight into how viral infection contributes to clinical features. This knowledge can help to find drugs, discover novel interventions, and design vaccines.

Chapter 2 identified the Epsilon variants that emerged in California, USA, with three unique spike mutations (S13I, W152C, and L452R). We observed that this variant rapidly increased in frequency in the population, which may indicate a higher transmissibility. The L452R mutations increase infectivity in vitro experiments. Interestingly, we observed reduced antibody

neutralization by this variant, suggesting its role in immune evasion. This observation emphasized more research needed on variants, specifically spike L452R mutation.

In **Chapter 3**, comparing protein-protein interaction (PPI) and functional genetics data of three coronaviruses (SARS-CoV, SARS-CoV2, and MERS-COV), we found shared and unique host proteins targeted by these viruses. We also investigated the cellular localization of viral proteins during overexpressed conditions and SARS-CoV2 infection in cells. We have found that protein-protein interactions provide insights into the mechanisms by which viruses manipulate cellular processes. Then we validated the role of SARS-CoV2 viral proteins interacting with host proteins (Tom70, IL17RA, PGES-2, and SigmaR1). Interestingly, out of these four targets, IL17RA was a unique functional interactor of SARS-CoV2 ORF8. We also found the clinical relevance of these targets using patients data. This suggests that PPI and genetic screen can be beneficial in finding repurposed drugs during a pandemic.

In **chapter 4**, we investigated the role of the ZIKV capsid on the host NMD pathway. We found that ZIKV infection and capsid expression disrupt the NMD pathway and downregulate nuclear UPF1, a crucial member of the NMD pathway. Members of NMD were found in the ZIKV capsid interactome, and capsid interacts with UPF1. Furthermore, ZIKV replication was enhanced when UPF1 was knockdown in NPC. This suggests ZIKV targets the host NMD pathway to enhance viral replication.

Chapter 5, continuing our previous observation, we further investigated the consequences of ZIKV mediated inhibition of the NMD pathway. We found that in ZIKV infected NPC, host transcripts interacting with UPF1 are reduced. Interestingly we saw an accumulation of host transcript in the nucleus during ZIKV infection, capsid expression, and UPF1 knockdown. We also found that knockdown of UPF1 leads to the trapping of *FREM2* mRNA, causing depletion of FREM2 protein. This depletion changes NPC differentiation potential. So, ZIKV mediated UPF1 downregulation leads to altered NPC differentiation through *FREM2* mRNA retention in the nucleus.

Understanding how viral proteins cause viral pathogenesis is vital for finding a cure or reduce viral pathogenesis. Host-viral protein-protein interaction gives insight into which proteins and pathways the virus targets. By interacting with host proteins, viruses hijack and dysregulate host pathways to benefit viral replication. This leads to disruption of normal cellular functions, which can lead to severe conditions even deaths. If we can find which pathways are dysregulated due to viral infection, we can reduce the severity and even deaths. Thus more research is needed

on viral pathogenesis to understand the molecular mechanism, thus helping to find the cure and preventing interventions.

Nederlandse samenvatting

Nederlandse samenvatting

Virale epidemieën en pandemieën leiden tot ernstige ziekte en sterfte. Nieuwe pandemische of epidemische virussen zijn afkomstig van dieren. Wereldwijde veranderingen, zoals urbanisatie, ontbossing, verlies van ecosystemen, en klimaatverandering, leiden tot onverwachte en onbedoelde contacten met wilde dieren. Zo ontstaan ook onbedoelde contacten met de virussen van deze dieren, wat kan leiden tot overdracht van dierlijke virussen naar mensen. Tegelijkertijd hebben we te maken met een groeiende wereldpopulatie en een toename in het aantal dicht bewoonde gebieden, waarin virussen zich gemakkelijk kunnen verspreiden. Zowel SARS-CoV2 als het ZIKV zijn afkomstig van dieren. Zonder enige bestaande immuniteit in de populatie kunnen deze virussen leiden tot substantiële ziekte en sterfte. Naarmate er meer infecties optreden in een populatie zullen er meer mutaties in deze virussen optreden, wat kan leiden tot meer aangepaste en overdraagbare virussen met een hogere virale 'fitness'. Virale eiwitten interacteren met gastheereiwitten en deze interacties zijn vaak onmisbaar voor efficiënte virale replicatie. Tegelijkertijd verstoren dit soort interacties vaak normale fysiologische processen, met ziekteverschijnselen als gevolg. Het onderzoeken van de rol van virale eiwitten in pathogenese kan helpen bij het vinden van geneesmiddelen of profylactische behandelmethoden. Om de rol van deze interacties te begrijpen begint men vaak met het besturen van individuele virale eiwitten en hoe deze interacteren met gastheereiwitten. Ontdekken hoe deze interacties leiden tot verstoorde cellulaire processen kan inzichten geven in de moleculaire mechanismen van virale pathogenese. In dit proefschrift wordt de rol van virale eiwitten in de overdraagbaarheid en pathogenese van SARS-CoV2 en ZIKV onderzocht.

H**oofdstuk 1** gaat over de geschiedenis, transmissieroute en moleculaire biologie van SARS-CoV2 en ZIKV. Deze virussen bevatten beiden een RNA-genoom en hebben een vergelijkbare levenscyclus. Echter, de transmissie routes van deze virussen en natuurlijke gastheer zijn compleet verschillend. De ernst van de ziekte en meest getroffen leeftijdsgroep zijn ook verschillend. SARS-CoV2 heeft zijn oorsprong in vleermuizen voordat het oversprong naar de mens. Het SARS-CoV2 'spike' eiwit is een structureel eiwit dat essentieel is voor het binnendringen van cellen. Dit eiwit heeft een relatief hoge mutatiesnelheid, wat kan leiden tot een verhoogde infectiviteit en overdraagbaarheid van het virus. In het afgelopen jaar zijn er verschillende virusvarianten gevonden met kenmerkende mutaties in het spike eiwit. Deze varianten kunnen mogelijk efficiënter cellen infecteren dan het oorspronkelijke virus. Verschillende spike mutaties lijken hiervoor verantwoordelijk te zijn. Tegelijkertijd, zijn er bepaalde mutaties in spike die zorgen dat ons immuunsysteem het virus niet meer goed kan

herkennen. Bijvoorbeeld, de L452R mutatie in de Epsilon SARS-CoV2 variant lijkt verantwoordelijk te zijn voor een zowel verhoogde virale transmissie en infectiviteit, als het ontwijken van het immune systeem.

ORF8 is een accessoir gen van SARS-CoV2 dat mogelijk betrokken is bij het onderdrukken van de immuunrespons. Proteomics experimenten en in vitro studies wijzen erop dat ORF8 interacteert met IL17RA en het cellulaire IL17RA systeem ontregelt. In COVID-19 patiënten zijn er signaalstoffen (cytokinen) gedetecteerd die geïnduceerd kunnen worden via de IL17RA signaaltransductieroute. Zowel het spike eiwit als ORF8 spelen een belangrijke rol in virale transmissie en pathogenese, maar er is nog veel te ontdekken over de biologie van deze eiwitten.

Van 2015-2016 zorgde de ZIKV epidemie voor een verhoogd aantal gevallen van foetale microcefalie, een hersenontwikkelingsstoornis waarbij het hoofd van de foetus klein blijft. Het capside eiwit van ZIKV is een structureel eiwit dat een cruciale rol speelt in de virale replicatiecyclus. Dit eiwit interacteert met een groot aantal cellulaire systemen en draagt het bij aan het ontwikkelen van hersenontwikkelingsstoornissen. De ZIKV capside bevindt zich in de celkern en beïnvloedt cruciale cellulaire systemen, zoals de NMD signaaltransductieroute. Deze signaaltransductieroute is een RNA surveillance mechanisme dat een rol speelt in de replicatiecyclus van RNA virussen, maar ook in de ontwikkeling van de hersenen. ZIKV capside interacteert met UPF1, een essentieel eiwit van de NMD signaaltransductieroute, en verlaagt zijn expressie. Deze verlaging is gunstig voor virale replicatie. Ook zorgt dit voor de accumulatie van *FREM2* mRNA in de celkern, wat resulteert in een verlaging van FREM2 eiwitniveaus. FREM eiwit is essentieel voor de ontwikkeling van de hersenen en de verlaging van de expressie van dit eiwit verandert de differentiatie van neuronale voorlopercellen (NPC). Dus, capside-gemedieerde manipulatie van de NMD signaaltransductieroute lijkt een rol te spelen in ZIKV pathogenese.

Het bestuderen van gastheer-virus interacties leidt soms tot de identificatie van nieuwe cellulaire doelwitten van virussen. Bergrijpen hoe deze interacties cellulaire signaaltransductieroutes verstoren kan vervolgens leiden tot nieuwe inzichten in virale pathogenese. Deze kennis kan ons helpen om nieuwe antivirale therapieën of vaccines te ontwerpen.

In **hoofdstuk 2** wordt in Californië in de Verenigde Staten, de Epsilon variant ontdekt. Deze variant bevat drie unieke mutaties, S13I, W152C en L452R. We observeerden dat deze

variant snel toenam in frequentie in de populatie, wat kan duiden op een hogere overdraagbaarheid. De L452R mutaties verhoogt infectiviteit in in-vitro experimenten. Ook wordt deze variant minder goed geneutraliseerd door antistoffen, wat suggereert dat deze mutatie een rol speelt bij het ontsnappen aan het immuunsysteem. Deze observatie laat zien dat er meer onderzoek naar varianten nodig is, in het bijzonder naar de spike L452R mutatie.

In **hoofdstuk 3** hebben wij middels eiwit-eiwit interactie studies en functionele genetica van drie coronavirussen (SARS-CoV, SARS-CoV2 en MERS-CoV) een aantal unieke en gedeelde gastheereiwitten gevonden die interacteren met deze drie virussen. We hebben ook de cellulaire lokalisatie van virale eiwitten in de context van overexpressie studies en SARS-CoV2 infectiestudies bestudeerd. We hebben gevonden dat eiwit-eiwit interacties inzichten geven in de mechanismen waarop virussen cellulaire processen manipuleren. Vervolgens hebben we de rol van deze SARS-CoV2 eiwit interacties met cellulaire eiwitten Tom70, IL17RA, PGES-2 en SigmaR1 gevalideerd. In het bijzonder hebben wij gevonden dat IL17RA een functionele interactie aangaat met SARS-CoV2 ORF8 en dat deze interactie mogelijk klinisch relevant is op basis van patiëntengegevens. Dit suggereert dat eiwit-eiwit interactie studies en genetische screens gebruikt kunnen worden voor het identificeren van bestaande medicijnen die wellicht gebruikt kunnen worden tegen een nieuw pandemisch virus.

In **hoofdstuk 4** hebben wij de rol van het ZIKV capside eiwit bestudeerd in de cellulaire NMD signaaltransductieroute. Wij hebben gevonden dat ZIKV infectie en capside expressie de NMD signaaltransductieroute verstoren en leiden tot een afname van nucleair UPF1, een cruciaal eiwit in de NMD signaaltransductieroute. Verschillende eiwitten in deze signaaltransductieroute zijn gevonden in het capside interactoom, en het capside eiwit interacteert met UPF1. Ook is ZIKV replicatie bevorderd na UPF1 ‘knockdown’ in NPC. Dit suggereert dat ZIKV de NMD signaaltransductieroute manipuleert om virale replicatie te bevorderen.

In **hoofdstuk 5** hebben wij voortgeborduurd op onze eerder observaties en de gevolgen van ZIKV-gemedieerde inhibitie van de NMD signaaltransductieroute bestudeerd. We demonstreren dat in ZIKV-geïnfecteerde NPC minder gastheertranscripten interacteren met UPF1. Ook zagen we een accumulatie van gastheertranscripten in de celkern tijdens ZIKV infectie, capside overexpressie en UPF1 ‘knockdown’. Verder hebben we gevonden dat UPF1 ‘knockdown’ ertoe leidt dat *FREM2* mRNA vast komt te zitten in de celkern, wat leidt tot een afname in FREM2 eiwitexpressie. Deze depletie verandert het differentiatiepotentieel van NPC.

Dus ZIKV-gemedieerde UPF1 afname leidt tot een veranderde NPC differentiatie door de retentie van *FREM2* mRNA in de celkern.

Voor de ontwikkeling van nieuwe medicijnen is het belangrijk om te begrijpen hoe virale eiwitten ziekte veroorzaken. Virus-gastheer interactiestudies geven inzichten in welke virale eiwitten in staat zijn om belangrijke signaaltransductieroutes te manipuleren. Door te interacteren met gastheereiwitten, kunnen virussen de cel manipuleren en geheel inrichten voor optimale virusreplicatie. Dit verstoort normale cellulair functies en kan leiden tot ernstige ziekte, en zelfs sterfte. Als we kunnen identificeren welke cellulaire processen verstoord zijn door een virus, dan kunnen we medicijnen ontwikkelen die de ernst van de ziekte kunnen verminderen. Kortom, meer onderzoek is nodig om de moleculair mechanismen achter virale pathogenese te begrijpen.

Curriculum Vitae

Mir Mubashir Khalid

Born November 1st, 1989, Gazipur, Bangladesh
mir.khalid@gladstone.ucsf.edu; khaliddrmc@gmail.com

Education

- Ph.D. Candidate, Research School of MGC (The Medical Genetics Centre South-West Netherlands), ErasmusMC, The Netherlands,
  Oct 2016-Present
  Advisors: Peter Verrijzer, PhD & Tokameh Mahmoudi, PhD

- Visiting Graduate Student, Gladstone Institutes, San Francisco,
  Oct 2016-Present
  Advisor: Melanie Ott, MD, PhD
  Area of Concentration: Molecular virology and Chromatin Biology.

- Master of Science (MSc) in Infection & Immunity (Research), Erasmus University Rotterdam, The Netherlands.
  Aug 2013-Aug 2015
  Area of Concentration: Infection & Immunobiology.

- Master of Science (MS) in Genetic Engineering & Biotechnology, University of Dhaka, Bangladesh.
  2011-2012
  Area of Concentration: Molecular Genetics & infections.

- Bachelor of Science (BS) in Genetic Engineering & Biotechnology, University of Dhaka, Bangladesh.
  2007-2011
  Area of Concentration: Molecular Cell Biology & Biochemistry.

Research Experience

- Graduate Student Researcher, Ott Lab, Gladstone Institute of Virology, Gladstone Institutes (UCSF), San Francisco, California, September 2016 – Present
  Research Advisor: Melanie Ott, MD, PhD.
  Project Title:
  - Generation of SARS-CoV2 Pseudotyped virus system for viral entry assay and Human Alveolar Organoids(HAO) infection model.
  - Understanding role of SARS-CoV2 Orf8 protein in IL17RA pathway.
  - Role of SARS-CoV2 Spike mutations in viral entry.
  - ZIKV Capsid's role on NMD (Nonsense-mediated mRNA decay) pathway.
  - ZIKV Capsid's role on Chromatin regulators and remodelers.

- Research Analyst, Mahmoudi Lab, Dept. of Biochemistry, ErasmusMC, Rotterdam, Sept 2015-Aug 2016
  Research Advisor: Tokameh Mahmoudi, PhD.

Project Title: Studying HBV infection and carcinogenesis in liver organoids.

- Research Intern, Mahmoudi Lab, Dept. of Biochemistry, ErasmusMC, Rotterdam, Jan 2014-August 2015
  Research Advisor: Tokameh Mahmoudi, PhD.
  Project Titles:
  - Liver Organoids: A new model system to study HBV infection
  - A haploid screen to identify host factors in HIV latency
  - Small molecule inhibitors of BAF as HIV Latency Reversal Agents (LRA).

- Research Officer & Fellow, Islam's Lab, Emerging Diseases and Immunbiology Lab, Centre for Food and Waterborne Diseases, icddrb, Dhaka. Nov 2012-Aug 2013
- Project Title:
  - Immunogenic polymorphisms in Guillain-Barré syndrome patients in Bangladesh
  - Prospective hospital-based surveillance on encephalitis in Bangladesh: clinical features, diagnostic imaging and etiology.
- Elective fellow, Islam's Lab, Emerging Diseases and Immunbiology Lab, Centre for Food and Waterborne Diseases, icddrb, Dhaka. July 2011-July 2012

Publications

1. Rapid assessment of SARS-CoV-2 evolved variants using virus-like particles. Syed AM,Taha TY, Tabata T, Chen IP, Ciling A, **Khalid MM**, Sreekumar B, Chen PY, Hayashi JM, Soczek KM, Melanie Ott M, Doudna JA. **bioRxiv. 2021.** doi: 10.1101/2021.08.05.455082 (Accepted in **Science)**
2. Application of human liver organoids as a patient-derived primary model for HBV infection and related hepatocellular carcinoma. De Crignis E, Hossain T, Romal S, Carofiglio F, Moulos P, **Khalid MM**, Rao S, Bazrafshan A, Verstegen MM, Pourfarzad F, Koutsothanassis C, Gehart H, Kan TW, Palstra RJ, Boucher C, IJzermans JN, Huch M, Boj SF, Vries R, Clevers H, van der Laan LJ, Hatzis P, Mahmoudi T. **Elife. 2021.** doi: 10.7554/eLife.60747.
3. Sentinel cells enable genetic detection of SARS-CoV-2 Spike protein. Weinberg ZY, Hilburger CE, Kim M, Cao L, **Khalid M**, Elmes S, Diwanji D, Hernandez E, Lopez J, Schaefer K, Smith AM, Zhou F; QCRG Structural Biology Consortium, Kumar RG, Ott M, Baker D, El-Samad H. **bioRxiv. 2021**. doi: 10.1101/2021.04.20.440678
4. Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant. Deng X*, Garcia-Knight MA*, **Khalid MM***, Servellita V*, Wang C*, Morris MK, Sotomayor-González A, Glasner DR, Reyes KR, Gliwa AS, Reddy NP, Martin CSS, Federman S, Cheng J, Balcerek J,Taylor J, Streithorst JA,Miller S, Sreekumar B, Chen PY, Schulze-Gahmen U, Taha TY, Hayashi JM, Siomoneau CR, Kumar GR, McMahon S, Lidsky PV, Xiao Y, Hemarajata P, Green NM, Espinosa A, Kath C, Haw M, Bell J, Hacker JK, Hanson C, Wadford DA, Anaya C, Ferguson D, Frankino PA, Shivram H, Lareau LF, Wyman SK, Ott M, Andino R, Chiu CY. **Cell. 2021.** doi: 10.1016/j.cell.2021.04.025.

5. Zika Virus Infection Prevents Host mRNA Nuclear Export by Disrupting UPF1 Function. Leon K, Flynn R*, **Khalid MM***, Fontaine KA, Nguyen T, Renuka Kumar GR, Simoneau CR, Tomar S, Jimenez-Morales D, Dunlap M, Kaye J, Shah PS, Finkbeiner S, Krogan NJ, Bertozzi C, Carette JE, Ott M. **bioRxiv. 2020.** doi:10.1101/2020.12.03.410837
6. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. Gordon DE, Hiatt J, Bouhaddou M, Rezelj VV, Ulferts S, Braberg H, Jureka AS, Obernier K, Guo JZ, Batra J, Kaake RM, Weckstein AR, Owens TW, Gupta M, Pourmal S, Titus EW, Cakir M, , …… **Khalid MM**, Vallejo-Gracia A, Fozouni P, Simoneau CR, ….. Ott M, Palmarini M, Shokat KM, García-Sastre A, Rassen JA, Grosse R, Rosenberg OS, Verba KA, Basler CF, Vignuzzi M, Peden AA, Beltrao P, Krogan NJ. **Science. 2020.** doi: 10.1126/science.abe9403.
7. Crosstalk between RNA Pol II C-Terminal Domain Acetylation and Phosphorylation via RPRD Proteins.Ali I, Ruiz DG, Ni Z, Johnson JR, Zhang H, Li PC, **Khalid MM**, Conrad RJ, Guo X, Min J, Greenblatt J, Jacobson M, Krogan NJ, Ott M. **Molecular Cell. 2019** doi:10.1016/j.molcel.2019.04.008
8. Toll-like receptor-4 299Gly allele is associated with Guillain-Barré syndrome in Bangladesh. Jahan I, Ahammad RU, **Khalid MM**, Rahman MI, Hayat S, Islam B, Mohammad QD, Islam Z. **Annals of Clinical and Translational Neurology. 2019.** doi:10.1002/acn3.744
9. The Cellular NMD Pathway Restricts Zika Virus Infection and Is Targeted by the Viral Capsid Protein.Fontaine KA, Leon KE, **Khalid MM**, Tomar S, Jimenez-Morales D, Dunlap M, Kaye JA, Shah PS, Finkbeiner S, Krogan NJ, Ott M. **MBio. 2018.** doi:10.1128/mBio.02126-18
10. CD1A and CD1E gene polymorphisms are not associated with susceptibility to Guillain-Barré syndrome in the Bangladeshi population. Rahman MI, Jahan I, **Khalid MM**, Jahan I, Ahammad RU, Nahar S, Islam Z. **J Neuroimmunol. 2018.** doi:10.1016/j.jneuroim.2017.11.013.
11. Tumor necrosis factor-alpha -863C/A polymorphism is associated with Guillain-Barré syndrome in Bangladesh.Jahan I, Ahammad RU, Farzana KS, **Khalid MM**, Islam MB, Rahman MI, Nahar S, Kabir Y, Mohmmad QD, Islam Z**. J Neuroimmunol. 2017.** doi:10.1016/j.jneuroim.2017.06.005.
12. Small Molecule Inhibitors of BAF; A Promising Family of Compounds in HIV-1 Latency Reversal.Stoszko M, De Crignis E, Rokx C, **Khalid MM**, Lungu C, Palstra RJ, Kan TW, Boucher C, Verbon A, Dykhuizen EC, Mahmoudi T. **EBioMedicine. 2015.** doi:10.1016/j.ebiom.2015.11.047.

***these authors contributed equally**

Employment History

- October 2016-present, Visiting Graduate Student, Gladstone Institutes, GIVI. P.I.: **Melanie Ott**
- November 2015–August,2016, Research Analysist, Erasmus MC, Dept. of Biochemistry. P.I.: **Tokameh Mahmoudi**
- November 2012–August, 2013, Research officer, icddrb, Emerging Diseases and Immunbiology Lab. P.I.: **Zhahirul Islam**

Leaderships

Organizer, Immunology and Virology Research in Progress Seminar Series, Gladstone Institutes, July 2019 – 2020; July 2021-22.

- Schedule and manage research progress talk by graduate students and postdocs
- Invite scientists outside home institute to present their talk in the seminar

Floor Ambassador, Graduate Student Organization, Gladstone Institutes, September 2019 – Present

- Serve as student council representative for GIVI graduate students, organize monthly student lunches and PI-Graduate student mentorship lunches/events, plan outreach & retreat events

Moderator, Bay Area COVID talk at Gladstone, 2021

Organizer of Symposium, Admin (website & Facebook Page), Committee Member, SURE (Student Union Research masters ErasmusMC), ErasmusMC, 2015-16

Congress Co-coordinator & member, Committee for 3rd Congress of YoungBB (Young Biotechnologist of Bangladesh), 2012

Fellowship & Scholarships

Fellowships/Scholarships	Year
I&I fund (Erasmus Vrienden Fonds)	2015
I&I fund for international activities	2015
Erasmus Infection & Immunity Grant	2013-2015
National Science and Information & Communication Technology (NSICT) fellowship	2012
Shahidullah Hall scholarship on BS result	2012
“Sumitomo Corporation, Japan” scholarship	2007-2010
Talent Pool Scholarship from the Ministry of Education, Bangladesh	2007-2010
Gazipur District Council Scholarship	2005-2008

PhD Portfolio

Courses & trainings

- Cystoscope and Network Biology: An Introduction Workshop, 2nd february 2020
- Introduction to Biology of Cancer, online non-credit course authorized by Johns Hopkins University and offered through Coursera, 2019
- iBiology Course: "Let's Experiment: A Guide for Scientists Working at the Bench"
- Grant Writing Series (NIH F32, F30/31), April, 2019
- Teledyne Photometrics course, 2 days course, April, 2019
- Responsible conduct of research program: Research Ethics: Plagiarism and Misconduct in Scientific Research, 2019
  Gladstone's ESL (English as Second Language) courses, 2018
- Illumina Bioinformatics Training, May, 2018
- Bloodborne Pathogen Training 2016-19
- Carcinogen Training, 2019
- Laboratory safety for researchers, 2016-2019
- Biosafety training, 2016
- Lab hazard assessment (LHAT) training, 2016

Workshops & Conferences

- BayViro Symposium, 2017-19
- Biohub Inter-lab Confab, 2019
- RNAseq Analysis using R, 2019
- Gladstone Scientific Retreat, 2019
- QBI/ Pasteur Symposium on Infectious Disease, 2019
- Biohub Ensemble Workshop, 2018
- Viral Immunology minicourse, 2018 (3 credit course)
- Training to tenure track workshop, Feb, 2017
- Mass Spectrometry workshops for biologists, Jan, 2017

Oral Presentations

- ZIKV E-dimer Vaccine: A New Hope, ImmunoX Journal Club, 2019
- Zika virus modulation of chromatin modifiers, GIVI Research In Progress Seminar, 2019
- Understanding ZIKV Capsid & host proteins interaction, GIVI Research In Progress Seminar, 2018
- Journal Club (Viral Immunology minicourse, 2018): IL28b genotype is associated with differential expression of intrahepatic interferon-stimulated genes in patients with chronic Hepatitis C. Urban TJ et al, Hepatology, 2010.
- Journal Club (Viral Immunology minicourse, 2018): Protection by immunoglobulin dual-affinity retargeting antibodies against Dengue virus. Brien JD et al, JVirol, 2013

Teaching Experience

SRTP (Summer Research Training Program-UCSF) Mentor, May 28th- July 31st, 2019

- Day to day supervising and teaching lab techniques
- Helped the student working on a project, presenting and writing a summer

Research Mentor, icddrb, 2012-11

- Day to day supervising and teaching lab techniques
- Guiding students to complete their MS thesis project

Posters

- Zika virus modulation of chromatin modifiers. Mir Khalid, Sakshi Tomar, Tom Nguyen and Melanie Ott. Gladstone Scientific Retreat, 2019
- Zika virus modulation of chromatin modifiers. Mir Khalid, Sakshi Tomar, Tom Nguyen and Melanie Ott. BayViro Symposium, 2019

Journal Club

- Viral Immunology minicourse, 2018
- ImmunoX Journal Club, 2019

Publications

1. Rapid assessment of SARS-CoV-2 evolved variants using virus-like particles. Syed AM,Taha TY, Tabata T, Chen IP, Ciling A, **Khalid MM**, Sreekumar B, Chen PY, Hayashi JM, Soczek KM, Melanie Ott M, Doudna JA. **bioRxiv. 2021.** doi:10.1101/2021.08.05.455082 (Accepted in **Science)**
2. Application of human liver organoids as a patient-derived primary model for HBV infection and related hepatocellular carcinoma. De Crignis E, Hossain T, Romal S, Carofiglio F, Moulos P, **Khalid MM**, Rao S, Bazrafshan A, Verstegen MM, Pourfarzad F, Koutsothanassis C, Gehart H, Kan TW, Palstra RJ, Boucher C, IJzermans JN, Huch M, Boj SF, Vries R, Clevers H, van der Laan LJ, Hatzis P, Mahmoudi T. **Elife. 2021.** doi: 10.7554/eLife.60747.
3. Sentinel cells enable genetic detection of SARS-CoV-2 Spike protein. Weinberg ZY, Hilburger CE, Kim M, Cao L, **Khalid M**, Elmes S, Diwanji D, Hernandez E, Lopez J, Schaefer K, Smith AM, Zhou F; QCRG Structural Biology Consortium, Kumar RG, Ott M, Baker D, El-Samad H. **bioRxiv. 2021**. doi: 10.1101/2021.04.20.440678.
4. Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant. Deng X*, Garcia-Knight MA*, **Khalid MM***, Servellita V*, Wang C*, Morris MK, Sotomayor-González A, Glasner DR, Reyes KR, Gliwa AS, Reddy NP, Martin CSS, Federman S, Cheng J, Balcerek J,Taylor J, Streithorst JA,Miller S, Sreekumar B, Chen PY, Schulze-Gahmen U, Taha TY, Hayashi JM, Siomoneau CR, Kumar GR, McMahon S, Lidsky PV, Xiao Y, Hemarajata P, Green NM, Espinosa A, Kath C, Haw M, Bell J, Hacker JK, Hanson C, Wadford DA, Anaya C, Ferguson D, Frankino PA, Shivram H, Lareau LF, Wyman SK, Ott M, Andino R, Chiu CY. **Cell. 2021.** doi: 10.1016/j.cell.2021.04.025.
5. Zika Virus Infection Prevents Host mRNA Nuclear Export by Disrupting UPF1 Function. Leon K, Flynn R*, **Khalid MM***, Fontaine KA, Nguyen T, Renuka Kumar GR, Simoneau CR, Tomar S, Jimenez-Morales D, Dunlap M, Kaye J, Shah PS, Finkbeiner S, Krogan NJ, Bertozzi C, Carette JE, Ott M. **bioRxiv. 2020.** doi:10.1101/2020.12.03.410837.
6. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. Gordon DE, Hiatt J, Bouhaddou M, Rezelj VV, Ulferts S, Braberg H, Jureka AS, Obernier K, Guo JZ, Batra J, Kaake RM, Weckstein AR, Owens TW, Gupta M, Pourmal S, Titus EW, Cakir M, , …… **Khalid MM**, Vallejo-Gracia A, Fozouni P, Simoneau CR, ….. Ott M, Palmarini M, Shokat KM, García-Sastre A, Rassen JA, Grosse R, Rosenberg OS, Verba KA, Basler CF, Vignuzzi M, Peden AA, Beltrao P, Krogan NJ. **Science. 2020.** doi: 10.1126/science.abe9403.
7. Crosstalk between RNA Pol II C-Terminal Domain Acetylation and Phosphorylation via RPRD Proteins.Ali I, Ruiz DG, Ni Z, Johnson JR, Zhang H, Li PC, **Khalid MM**, Conrad RJ, Guo X, Min J, Greenblatt J, Jacobson M, Krogan NJ, Ott M. **Molecular Cell. 2019** doi:10.1016/j.molcel.2019.04.008.
8. Toll-like receptor-4 299Gly allele is associated with Guillain-Barré syndrome in Bangladesh. Jahan I, Ahammad RU, **Khalid MM**, Rahman MI, Hayat S, Islam B, Mohammad QD, Islam Z. **Annals of Clinical and Translational Neurology. 2019.** doi:10.1002/acn3.744.
9. The Cellular NMD Pathway Restricts Zika Virus Infection and Is Targeted by the Viral Capsid Protein.Fontaine KA, Leon KE, **Khalid MM**, Tomar S, Jimenez-Morales D, Dunlap M, Kaye JA, Shah PS, Finkbeiner S, Krogan NJ, Ott M. **MBio. 2018.** doi:10.1128/mBio.02126-18.

10. CD1A and CD1E gene polymorphisms are not associated with susceptibility to Guillain-Barré syndrome in the Bangladeshi population. Rahman MI, Jahan I, **Khalid MM**, Jahan I, Ahammad RU, Nahar S, Islam Z. **J Neuroimmunol. 2018.** doi:10.1016/j.jneuroim.2017.11.013.
11. Tumor necrosis factor-alpha -863C/A polymorphism is associated with Guillain-Barré syndrome in Bangladesh.Jahan I, Ahammad RU, Farzana KS, **Khalid MM**, Islam MB, Rahman MI, Nahar S, Kabir Y, Mohmmad QD, Islam Z**. J Neuroimmunol. 2017.** doi:10.1016/j.jneuroim.2017.06.005.
12. Small Molecule Inhibitors of BAF; A Promising Family of Compounds in HIV-1 Latency Reversal.Stoszko M, De Crignis E, Rokx C, **Khalid MM**, Lungu C, Palstra RJ, Kan TW, Boucher C, Verbon A, Dykhuizen EC, Mahmoudi T. **EBioMedicine. 2015.** doi:10.1016/j.ebiom.2015.11.047.

***these authors contributed equally**

Acknowledgments

Ph.D. is a journey that I have cherished and wanted since I found my research interest in biomedical science. Scientists are both superheroes and detectives who spend their lives finding answers to questions regarding the natural world, which leads to saving lives and the world. Curiosity and scientific observation are essential to ask a question that will lead to a scientific explanation. My interest in science led me to work in three different countries and three prestigious research institutes. It was a fantastic experience working on the research in this current thesis. As a person who loves science and research, I enjoyed every bit of it. I love staying up late for experiments; I get excited when I find new observations or data supporting my hypothesis. Working with intelligent scientists from different parts of the world improved my understanding of science. I learned a lot, corrected mistakes, improved experimental designs and analysis. This Ph.D. research has contributions, inspiration, and help from amazing and supportive people. I cannot imagine completing all these without them.

First, I want to thank **Prof.dr. Melanie Ott** for supervising me. She is a caring and brilliant scientist who, over all these years, believed, supported, and helped me to pursue my research. This Ph.D. would not be possible without the help and support from her. Besides research, Melanie is a very supportive and charming person who always looks after her mentees and employees. I learned a lot from you, and that shaped how I think scientifically. You taught me how to think and observe critically. Your collaborative efforts changed my view about science. After the COVID-19 pandemic was declared, you established the BSL3 facility at Gladstone Institutes. Your strong leadership and continued efforts paved the way for SARS-CoV2 research. At the same time, you made sure everyone was safe and happy. Thank you for your patience and support. Thank you very much for being a unique and wonderful supervisor. I am grateful to have you as my mentor.

I am very grateful to **Prof.dr. Tokameh Mahmoudi,** who regularly monitored my Ph.D. timeline and goals. Since I could not travel to the Netherlands or other countries due to VISA issues, she ensured all the administrative and academic matters in the Netherlands were taken care of. At the same time, every time I talked with she gave good advice, suggestions, and support. Thank you for your continued support and help. You are a fantastic scientist and mentor. I had a great time in your lab during my MSc thesis and later as a research analyst. You gave me the courage and support all these years. Thank you for giving me advice and guidance for my Ph.D. work.

Thank you, **Prof.dr. Peter Verrijzer,** for being the promoter of my Ph.D. In the Netherlands, you gave me the support and courage to start my Ph.D. I still remember discussing doing my Ph.D. research in the USA with you while being a Ph.D. candidate in the Netherlands. You told me not to worry about administrative issues in the Netherlands and go to the USA to do good research. That gave me a lot of confidence and courage to start this journey. Thank you for the support and help.

I want to thank **Marike van Geest** for taking care of all the administrative work and arranging my remote access to ErasmusMC. Each time I had an inquiry, you solved it and answered quickly. There are probably more than 200 emails I had exchanged with her, and at the same time, she reminded necessary steps when needed. A truly amazing and supportive person that any Ph.D. student needs. I thank you for answering all the questions, making sure I am well informed. Without you, I cannot imagine graduating. Thanks to my thesis committee (**Prof.dr. Eric Verdin, Dr. Bart Haagmans,** and **Dr. Nadia Roan**) for reviewing my thesis and approving my promotion. I am very grateful to them. I am grateful to **Prof.dr. Eric C.M. van Gorp** and **Dr. Priya S. Shah** for accepting to participate in my thesis defense committee.

I want to thank all members of **Ott lab** who are very supportive. Working in the lab with my colleagues gave me a feeling of a family. Thank you for being excellent and friendly. Thank you for tolerating my pranks and jokes. I first want to thank my bay members, who supported and helped throughout my Ph.D. journey. **Irene**, you are a unique, funny, sweet, and very supportive friend. Over the year, we did many pranks, amusing poses, and funny gigs. Half of your phone gallery is probably random selfies Frank, and I take whenever your phone is left unattended. Also, I send random emails or texts to your friends from your device saying, 'Khalid is a great scientist' to make them confused. Thank you for being a fantastic friend and colleague. You give excellent advice and criticism to experiments and results, which make data more trustworthy. Thank you for all the feedback and corrections on my thesis. At the same time, thank you for tolerating me as a bay-mate. Thank you for all the support all these years.

I want to thank **Parinaz,** who sat next to me in the lab for almost four years. Thank you for tolerating me and my sarcastic comment. I enjoyed our late-night work and chatting. Your help in getting me on some UCSF courses was very helpful. At the same time, you are my savior and redeemer. This is why I have your picture as a deity to support my research. You bring happiness and energy to the workplace. I enjoyed working with you and having you as a bay-mate. Also, I am always eager to listen to stories of your beautiful nieces Ava and Nilu (little munchkins).

Albert, my friend, you are a sweet, caring, and loving person. You are the longest and the earliest lab member who was always with me even we moved our bay twice. I enjoyed your excitement about board games and any fun activities, and I hope to meet you when I go to Europe. Thank you for coming to our house and being a fantastic friend to Asha.

2021 brought new members to my bay. I want to thank **Frank,** my buddy and bay-leader is a fantastic scientist and fun to work with. Thank you very much for checking and reading my thesis thoroughly. You and I are the two dimers of ORF8. You are a fun-loving guy! It is fun to work with you and play computer games with you. Keep skate-boarding. I love our discussion on science and other issues. Hanging out with you is always fun, and you are full of energy. And your cat Kelby is so cute and friendly. He is a fantastic host whenever we hang out at your place.

Tessa, you are unique and funny. Your athleticism fascinates me. Keep doing vaulting. Whenever I see you, I feel like you are full of energy and fun. Thank you for your support and for being amazing. I enjoy our little chat and talking about your hobby. Thanks to **Rahul,** who joined our bay very recently. Working with you is fun, and I am excited about the current projects we are working on. Thank you for your support and for being an amazing colleague.

From 2016 and still, I worked with a lot of amazing scientists in the Ott lab. I want to thank **Sarah;** I enjoyed the times we spent talking about all science topics and non-science topics. Thank you for helping me, not only with the SARS-CoV2 project but also with other times. Also, I want to thank **Rory,** Sarah's dog, and cookie monster. Rory, you big fluffy cookie monster, I am going to miss you a lot. Whenever I see you, and you jump at me, it makes me feel relaxed. You would be the best medicine to take away anxiety and stress.. Thanks to **Kris,** who was also interested in ZIKV research. You taught me how to propagate the ZIKV virus, Immunofluorescence, and NPC works. I am excited to finish our last work, which is chapter 5 of my thesis. I thank you for being awesome and feeding me fantastic foods. When I joined Ott lab, I worked with **Krystal** and **Nathan**. I thank both of them for guiding and helping me. I valued your guidance, patience, and mentoring. Working with you gave me the confidence to work on my projects, and I was thrilled that I was able to contribute to your publications. I am thankful to both of you. I want to thank **Kathrin** for being an awesome friend. We had so much fun in SF. Albert, you and I did lots of things together. You are a wonderful colleague and friend. I love your suggestions on my projects. Hanging out with you was always fun. When I visit Europe, I am going to see you.

Chia-Lin, without you in the lab, everything will stop. During the COVID-19 pandemic, when the BSL2 lab supply was scarce, I appreciated you contacted so many vendors to get all those supplies. You are very organized and helpful. Thank you for helping me with researches and support. I appreciated your advice and support. For any lab problems, you are always there to help. Also, thank you for introducing me to Mei-Zhong, the cute little dog. I want to thank **Renu** for guiding and helping on SARS-CoV2 projects. Thank you for taking the time to listen to my problems and giving me good advice. Your helpful direction and support enabled me to contribute much SARS-CoV2 research. Talking with you is always very relaxing, and I am looking forward to working more with you.

I want to thank 'pseudotyped virus assay' team aka 'Boo Boo Gang'. **Taha**, thank you for helping and advising me on experiments. You are a very knowledgeable and helpful person. Your knowledge of any technique is extraordinary. You are very good at any sport, and it is hard to beat you in sports. You also keep pushing me to graduate. Thank you also for giving me feedback on my thesis. **Bharath,** you are a fantastic person. Thank you for helping me pseudotyped virus assay and tolerating my jokes. I like hanging out with you and talking to you. Thank you for your support and amazing friend. Special shout out to **Jesse** as our HAO organoids provider. Thank you for helping me with my projects. You are a fantastic and hardworking scientist. I love hanging out with you and playing games. Thank you very much for helping me and working pseudotyped assay project. Thank you for also supplying me with organoids for my experiments. **Jen,** thank you very much for being a fantastic colleague and support. Thank you not only for the pseudotyped virus project but also for the Orf8 project. You are also one of the earliest bay-mate. We collaborated before and still collaborating. Thank you for introducing me to Lucy, Goose and Sheva; allare beautiful dogs.

My first bay-mate, **Sakshi,** I thank you for helping with ZIKV research. Working with you was fun. Thank you very much for helping me with my projects. I am grateful for all the help you gave. **Camille,** I am thankful for your suggestions and advice on my projects. Your tips on figures and data presentations made my presenting skill better. I am also happy that we worked on a project together. Thank you for all the support and collaboration. I want to thank **Gili;** you are full of fun and energy. Inside BSL2, we chat a lot, and I love our discussion about research and life. Thank you for all the help and fun times in the lab. Your dog, Humi, is so cute. His eyes are so pretty, and I am hoping to meet him one day. **Danielle,** thank you for tolerating my pranks and jokes. I love our late evening chatting in my bay. Half of my photoshopping pranks are your pictures (provided by you). I will never forget that we arranged a funeral for your bench since you were

working from home during the pandemic. When you came and saw it, you loved it. Thank you for being an excellent colleague.

I want to thank **Ursula** for your suggestion and help. Your insight into structural biology is amazing. Thank you for giving me inputs on my research. **Daniella,** you helped me many times with suggestions and advice. Thank you for helping me on different occasions from the beginning of my Ph.D. **Takako,** thank you for all the support and help on the collaborative work we did for a collaborator. Your cats, Kiki and Koko are adorable. Thank you for also being a supportive person. Thank you **Jeffrey**, you are so helpful and sweet. You would send me ZIKV articles and discuss experiments with me. Your suggestions and ideas were very helpful. Thank you for being awesome. Thanks to **Stephanie,** all the questions I asked about TECAN. Any questions I have, you are always there to help me. Thanks to **Veronica** and **Lauren** for taking care of administrative works for the lab. I want to thank you for all your support and help.

Joao and Dawlyn, thank you for your help and support. thank you for all the suggestions and help. **Raina**, the SRTP student who worked with me, I am thankful for your time and help. It was an outstanding experience mentoring you. The 'sloth' (my favorite animal) you made for me with bead stringing inspires me to do science. You joined a graduate program at UC, Santa Cruz. An amazing scientist is training and cannot wait when you graduate.

Ansu(Philip) and **Ibs(Ibraheem)**, you guys are the best. You helped me in the lab, and we are good friends. You are my brothers from another mothers. I am very proud of you, and your success. I am always excited to meet you. **Ryan and Olivia,** thank you for being amazing colleagues and for your supports.

I want to thank **Doudna lab**. **Syed**, you are an amazing person to work with, and you are always there to help. Thank you for all the support and help. **Karen** and **I-Li,** thank you for your support.I also want to thank 5th-floor colleagues. **Sonali,** thank you for being a fantastic friend, scientist sister, and for your support. I am very proud of you. **Mauricio** and **Roland**, thank you for your help on different occasions. **Joe**, thank you for being awesome and solving the Rubik's cube with me. We also did fantastic research together. I want to thank **Shahazada** bhai and **Samee bhai**, two other Bangladeshis who later left Gladstone Institutes. You helped me going through lots of stuff and helped me to get started living in the USA. **Hesham, Devin, Irene(Lew), Frank, Ujjwal, Arjun, Ravi, Arturo, Nick, Grace (Guorui), Jason,** and **Xiaoyu** for support discussion and help. I also want to thank **Xing, Austin, Swetansu, Alexi** from the other floors of Gladstone Institutes.

Thanks to **Hannah** from Pillai lab for helping me. **QCRG consortium** and **Krogan lab,** thank you for helping with reagents, knowledge, and support. I also want to thank Gladstone Institutes **HR** and **Purchasing department.** I want to thank **Sudha,** who is very supportive and helpful. I also want to thank **Gladstone's Graduate Organization** (Gladstone GO). I was floor ambassador or for two years. Working with Gladstone GO was an amazing experience. Thanks to **UCSF improv, JKA Karate Dojo(Mission Bay)**, and **Gladstone Soccer group,** for making my life entertaining and enjoyable. **Robert** (Bob) **Mahley** and **Linda Mahley**, thank you for all the fantastic talks and teaching.

Mateusz, **Enrico,** and **Michael** from Mahmoudi lab thank you for answering my graduation questions. I appreciate your support in helping me guiding on this thesis. I want to thank my friends in the Netherlands. **Chinmoy** bhai, thank you for giving me moral support and inspiring me. I like our discussion on many issues. **Mart**, thank you for visiting me in SF and being fantastic friends. Thank you for translating my summary into Dutch. Your input on the summary was extremely helpful. We talked a lot about research. It is always fun to talk with you. It was fun working and hanging out with you in the Netherlands. Hopefully, we will meet again and will do lots of fun stuff. **Rory(de Vries) and Stalin** thank you for visiting me in the USA and being a fantastic friend.

I thank my country, **Bangladesh,** people, and all the freedom fighters. My heart always belongs to Bangladesh, my motherland. I want to thank my three best friends in Bangladesh. **Faisal(Miraj), Sajid,** and **Nayim** for being amazing friends. I enjoyed all the talking and time we spent on friendship. Even now, we are living far away, and we still talk-chat on social media. I want to thank **my friends and fans/followers** on social media and my blogs. Thank you all for your continued support of my science blogs, and your support inspires me to continue my research. I also want to thank **Aubhishekh**, **Marwah,** and **Tareque**, my Bangladeshi friends who also live/lived in California and made many good moments at my place.

My parents, Amma(Mom) and Abba(dad), you both inspired, supported, and helped me all my life. **Amma,** there is no word to describe how an amazing person you are. Your sacrifice and struggle to educate all her children are outstanding. You always put us at a higher priority than anything in your life. **Abba,** you have always called me a scientist from my childhood. Your honesty and boldness were my mottoes for life. Your health deteriorated from the beginning of my Ph.D. and worsened severely at the end of my thesis. But you told me to finish my Ph.D. and not worry about you. Both of you taught me to chase my aim and dream. I cannot imagine being

in this position without you. I want to thank my **siblings,** who are the best and supportive. They take care of all the issues in Bangladesh for me.

I want to thank **my in-law family** for their continued support and help. My **father-in-Law** (Baba) and **mother-in-law** (Ma) are very supportive. Almost every week, we talk. It is very relaxing and fun to talk with you both. I want to thank my only brother-in-law, (Atiq bhai) and sister-in-law (Joyee vabi) for their support. I want to thank all my nephews and nieces (Sara, Saad, Adiba, Ahnaf, and Ariya) for asking me to come to their homes and sometimes sending planes(!). I miss you all and am excited to see Ahnaf and Ariya on my next visit to Bangladesh.

And last but not least, I want to thank the love of my life, my friend, my dear wife, **Habiba (Asha),** for being supportive. You never complained about me working late. Whenever I was stressed, you comforted me and inspired me. I am grateful and fortunate to have you as my life partner. It will be challenging to complete my Ph.D. without your help. Thank you for your support and unconditional love. Thank you for loving my cooking and my jokes. I am also very excited and proud that you started a master's degree in Public Health at Berkeley. Love you!

If there is someone I forgot to thank, I apologize from the bottom of my heart. I want to thank everyone who supported and helped me with this journey. I am happy and grateful for all the supports and help I got, and I appreciate all the contribution and help I got.

Thesis title: Exploring the molecular mechanisms of SARS-CoV2 and ZIKV pathogenesis

Propositions:

1. Epsilon variant (B.1.427/B.1.429) has higher infectivity, increased transmission, and decreased antibody neutralization. (This thesis)
2. SARS-CoV2 spike protein mutations like L452R can increase the fitness of the virus with increased infectivity and immune evasion properties. (This thesis)
3. Understanding viral-host protein-protein interaction from pan-coronaviral proteomics reveals unique and shared mechanisms that virus employs to enhance viral replication and immune evasion. (This thesis)
4. Accessory protein ORF8 is a mysterious protein with immunomodulatory functions. (This thesis)
5. Disrupting the RNA surveillance pathway by ZIKV capsid protein helps viral replication by downregulating host defense against RNA viruses. (This thesis)
6. Current evidence strongly supports the natural origin of SARS-CoV2 than suspicions on lab-leak and lab-made origins. (Andersen, K.G., Rambaut, A., Lipkin, W.I. et al. Nat Med (2020))
7. It was a great victory for science to eradicate smallpox from the earth, saving millions of lives. (Reingold A. Smallpox—The Death of a Disease: The Inside Story of Eradicating a Worldwide Killer: By D. A. Henderson. Am J Epidemiol. (2010))
8. Viruses are the magnificent entity that can infect every living organism on earth and influences their life. (A Planet of Viruses: Second Edition (2015) by Carl Zimmer)
9. Distinct compositions of the BAF complex determine the fate and differentiation of embryonic stem cells toward mammalian neural development. (Son EY, Crabtree GR. Am J Med Genet C Semin Med Genet. 2014)
10. Mass extinction shaped the big-brained mammal evolution. (Smaers JB, Rothman RS, Hudson DR, et al. Sci Adv. 2021)
11. "Being strong isn’t just about having power or move, it about one’s spirit."- Roronoa Zoro, 'One Piece' by Eiichiro Oda.

Thesis title: Exploring the molecular mechanisms of SARS-CoV2 and ZIKV pathogenesis

Propositions:

1. Epsilon variant (B.1.427/B.1.429) has higher infectivity, increased transmission, and decreased antibody neutralization. (This thesis)
2. SARS-CoV2 spike protein mutations like L452R can increase the fitness of the virus with increased infectivity and immune evasion properties. (This thesis)
3. Understanding viral-host protein-protein interaction from pan-coronaviral proteomics reveals unique and shared mechanisms that virus employs to enhance viral replication and immune evasion. (This thesis)
4. Accessory protein ORF8 is a mysterious protein with immunomodulatory functions. (This thesis)
5. Disrupting the RNA surveillance pathway by ZIKV capsid protein helps viral replication by downregulating host defense against RNA viruses. (This thesis)
6. Current evidence strongly supports the natural origin of SARS-CoV2 than suspicions on lab-leak and lab-made origins. (Andersen, K.G., Rambaut, A., Lipkin, W.I. et al. Nat Med (2020))
7. It was a great victory for science to eradicate smallpox from the earth, saving millions of lives. (Reingold A. Smallpox—The Death of a Disease: The Inside Story of Eradicating a Worldwide Killer: By D. A. Henderson. Am J Epidemiol. (2010))
8. Viruses are the magnificent entity that can infect every living organism on earth and influences their life. (A Planet of Viruses: Second Edition (2015) by Carl Zimmer)
9. Distinct compositions of the BAF complex determine the fate and differentiation of embryonic stem cells toward mammalian neural development. (Son EY, Crabtree GR. Am J Med Genet C Semin Med Genet. 2014)
10. Mass extinction shaped the big-brained mammal evolution. (Smaers JB, Rothman RS, Hudson DR, et al. Sci Adv. 2021)
11. "Being strong isn’t just about having power or move, it about one’s spirit."- Roronoa Zoro, 'One Piece' by Eiichiro Oda.